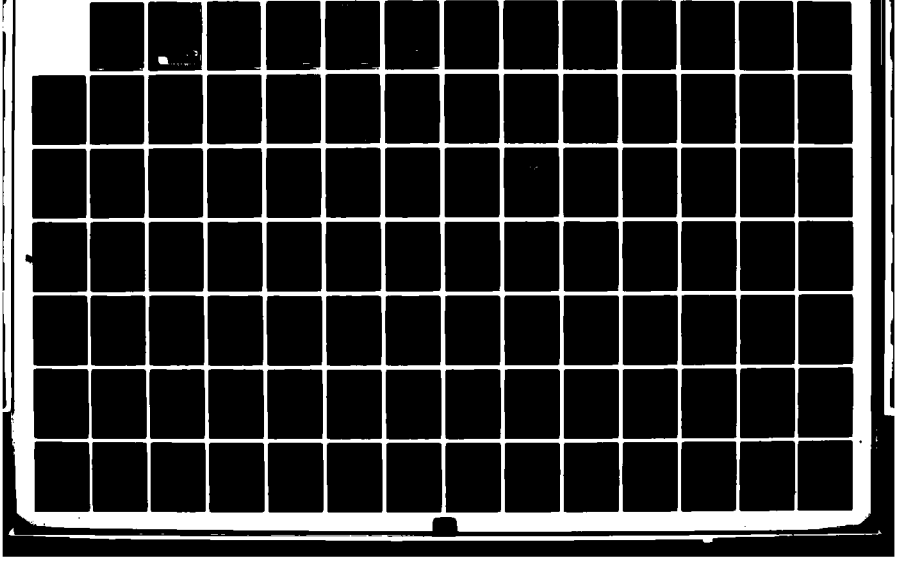
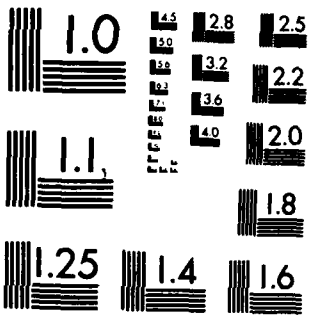
AD-A128 283 EVALUATION OF INFRARED TARGET DISCRIMINATION ALGORITHMS 1/2
(U) INSTITUTE FOR DEFENSE ANALYSES ALEXANDRIA VA
SCIENCE AND TECHNOLOGY DIV I W KAY APR 83 IDA-P-1714
UNCLASSIFIED IDA/HQ-83-25281 F/G 12/1 NL





(12)

IDA PAPER P-1714

EVALUATION OF INFRARED TARGET DISCRIMINATION ALGORITHMS

Irvin W. Kay

April 1983

INSTITUTE FOR DEFENSE ANALYSES
SCIENCE AND TECHNOLOGY DIVISION

The work reported in this document was conducted under IDA's Independent Research Program. Its publication does not imply endorsement by the Department of Defense or any other government agency, nor should the contents be construed as reflecting the official position of any government agency.

This document is unclassified and suitable for public release.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A128 283	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Evaluation of Infrared Target Discrimination Algorithms	5. TYPE OF REPORT & PERIOD COVERED FINAL May 1982-February 1983
	6. PERFORMING ORG. REPORT NUMBER IDA Paper P-1714
7. AUTHOR(s) Irvin W. Kay	8. CONTRACT OR GRANT NUMBER(s) IDA Independent Research
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Defense Analyses 1801 N. Beauregard Street Alexandria, VA 22311	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE April 1983
	13. NUMBER OF PAGES 131
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

This document is unclassified and suitable for public release.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

None

18. SUPPLEMENTARY NOTES

N/A

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Infrared Target Discrimination Algorithms, Point Targets, Clutter Rejection, Tracking Algorithms for Target Discrimination, Background Normalization, Multi-Color Target Discrimination, Terrain Background Models, Linear Filters for Target Discrimination

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper is concerned with the evaluation of algorithms used by passive infrared sensors to discriminate between signals due to target sources and those due to background clutter. The discussion is essentially restricted to the case of point targets.

The goal is to obtain a rough estimate of performance against minimum standards. For this purpose the analysis assumes a simple mathematical model for the background clutter

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (Continued)

distribution: namely, that it is multivariate Gaussian over the spatial and spectral data channels provided by the sensor. The paper also discusses experimental evidence for and against such a model, as well as certain more explicit statistical models that have been proposed for the spatial distribution of clutter.

Other topics discussed are CFAR optimum processing, linear filters, the effect of using ratios of spectral components for processing in multi-color systems rather than the components, themselves, and background normalization. Also discussed is the relationship between the effectiveness of tracking algorithms and the preliminary screening of targets by CFAR detection algorithms.

IDA PAPER P-1714

EVALUATION OF INFRARED TARGET DISCRIMINATION ALGORITHMS

Irvin W. Kay

April 1983

INSTITUTE FOR DEFENSE ANALYSES
SCIENCE AND TECHNOLOGY DIVISION
1801 N. Beauregard Street, Alexandria, Virginia 22311

IDA Independent Research Program

ACKNOWLEDGMENTS

The author wishes to acknowledge with gratitude the help contributed by Dr. Hans Wolfhard, Mr. Lucien Biberman, and Dr. Mark Burns, at IDA, in the form of critical reviews and suggestions of a substantive nature, as well as their willingness to discuss on many occasions questions related to various topics covered in the paper.

ABSTRACT

This paper is concerned with the evaluation of algorithms used by passive infrared sensors to discriminate between signals due to target sources and those due to background clutter. The discussion is essentially restricted to the case of point targets.

The goal is to obtain a rough estimate of performance against minimum standards. For this purpose the analysis assumes a simple mathematical model for the background clutter distribution: namely, that it is multivariate Gaussian over the spatial and spectral data channels provided by the sensor. The paper also discusses experimental evidence for and against such a model, as well as certain more explicit statistical models that have been proposed for the spatial distribution of clutter.

Other topics discussed are CFAR optimum processing, linear filters, the effect of using ratios of spectral components for processing in multi-color systems rather than the components, themselves, and background normalization. Also discussed is the relationship between the effectiveness of tracking algorithms and the preliminary screening of targets by CFAR detection algorithms.

CONTENTS

EXECUTIVE SUMMARY

Separating targets from their backgrounds is a signal processing problem that is a major concern to infrared sensors. This paper reviews several of the approaches that are now under serious consideration for use by infrared surveillance systems to deal with the problem--particularly for the case of point targets.

In the course of the review certain observations and conclusions scattered throughout the text may have more value for those who have a general interest in evaluating alternative approaches to the infrared target discrimination problem than other parts of the text. The parts that contain the supporting analysis must, of necessity, be somewhat drawn out and mathematically formal in order to provide the rigor needed to make a hard comparison between methods, or to disprove a common assumption. Thus some, perhaps most, of the material in this paper consists of technical detail that, undoubtedly, will be largely ignored by many readers whose interest in signal processing theory is only peripheral.

Therefore, the following summary is presented in an attempt to gather together the essence of this paper in the hope that it may, thereby, be rendered more accessible to the reader whose interests are less specialized. Each item is headlined and annotated for easy reference to the pertinent analysis or discussion contained in the main body of the paper.

EXPERIMENTAL SUPPORT FOR STATISTICAL MODELS OF TERRAIN BACKGROUND

Empirical evidence indicates that infrared radiance from natural terrain, such as a forest or a desert, is, to a good

approximation, normally distributed for a variety of wavelength bands in both the solar and thermal regions of the spectrum. This is less true of scenes that have been affected in some way by protracted human intervention, e.g., farm land, proving grounds, large cities. In general, the approximation is better at night than during the day.

On the other hand, empirical evidence does not support certain theoretical models that have been proposed for the statistical spatial distribution of terrain background radiance. Specifically, the data are inconsistent with the so-called two-dimensional Markoff process distributions that are characterized by exponential correlation functions. In fact, some versions of this type of model are not even theoretically self-consistent. (More detailed discussions of these matters and supporting analyses appear in Chapter II, Section D.)

SUB-OPTIMAL NATURE OF LINEAR FILTERS

For Constant False Alarm Rate (CFAR) detection of point targets, linear filters are sub-optimal in general. The linear filter that, in theory, maximizes the signal-to-noise ratio for a background whose spatial distribution is statistically homogeneous is a limiting case that the true, nonlinear, optimal filter would approach if the temperature of the target were very large compared to that of the background and the size of the target were small compared to the Instantaneous Field Of View (IFOV) of a single infrared detector. (The analysis supporting these conclusions appears in Chapter III, Section B.)

THE VALUE AND LIMITATIONS OF TRACKING ALGORITHMS

- Target discrimination algorithms are of two types: those whose purpose is clutter rejection and those, referred to as tracking algorithms, that distinguish targets by their characteristic trajectories. Most infrared systems use both types.

Tracking algorithms, which are a form of Moving Target Indication (MTI) technique are, in principle, the only hope for achieving the very low false alarm rates that are typically required for the detection of point targets by Infrared Search and Tracking (IRST) system specifications. Nevertheless, it is also necessary for this purpose to provide preliminary clutter rejection means, such as spatial filtering and adaptive thresholding in one or more spectral channels, to reduce the number of false detections before invoking tracking algorithms.

System designers often attribute the reason for requiring a preliminary clutter rejection process to limitations the available computer capacity, i.e., memory size and comp er speed. This would seem to imply that technological adv es, e.g., the introduction of VHSIC and VLSIC, will eventua make such a procedure unnecessary.

However, the requirement is actually independent of computer capacity. That is, tracking algorithms will work only if, initially, the expected number of false detections is below a certain critical value. Moreover, the effectiveness of a tracking algorithm is extremely sensitive to errors unless the a priori false detection probability can be made small by those other, preliminary, signal processing techniques. (The analysis supporting these conclusions appears in Chapter III, Section C.)

THEORETICAL IMPLICATIONS AND POSSIBLE IMPROVEMENT OF BACKGROUND NORMALIZATION

A common form of adaptive thresholding, sometimes known as "background normalization," which is an averaging process implemented with a two-dimensional linear filter, is equivalent to a least-square-error fit of a linear function to data obtained from measurements of the background radiance distribution. It follows that the next order improvement would be a quadratic

least-square-error fit. The quadratic fit can also be accomplished by means of an averaging (weighted in this case) process that is implemented with a two-dimensional linear filter. (The derivation of these results appears in Chapter IV, Section B.)

THE FALSE ALARM PENALTY IMPOSED BY THE USE OF SPECTRAL COMPONENT RATIOS

For multi-color or spectral discrimination systems it is sometimes the practice to work with ratios of spectral components rather than the components themselves. For example, a two-color system with radiance measurements J_1 and J_2 in the two spectral bands would use a one-dimensional target discrimination algorithm operating on the ratio J_1/J_2 rather than a two-dimensional algorithm operating on the pair J_1, J_2. This usually results in significantly higher false alarm rates, sometimes by several orders of magnitude, than would be generated by the equivalent two-dimensional process. (The proof of these conclusions appears in Chapter IV, Section C.)

I. INTRODUCTION

This paper is concerned with algorithms used by passive infrared (IR) sensors to discriminate between signals due to target sources and those due to background clutter. The primary objective is to formulate a simple methodology for evaluating such algorithms.

The goal has been to develop an evaluation procedure that is easier to implement and is less specific than a detailed computer simulation, which is the usual approach to this objective. The purpose here is not to supplant computer simulation as a means of evaluating a signal processor's logic design. Rather, it is to provide an analytical tool that can be used for a rough, preliminary assessment of the feasibility or the potential of different processing schemes.

The scope of this paper is essentially restricted to the case of non-imaging systems, i.e., those, such as the infrared search and tracking system (IRST), for which targets behave as point sources under ordinary operating conditions.* The point target assumption implies that discrimination algorithms must be of an abstract nature, relying upon certain target and background signatures that are not associated with easily identified geometric attributes, such as size and shape, that would be available to an imaging system. However, signatures may be derived from any combination of spectral and temporal, as well as certain limited spatial, properties of targets and backgrounds.

*The term "non imaging" seems appropriate in the present context even if the system in question is capable of providing an image of the background (although not of the target) as long as it does not, in fact, make use of such an image for target discrimination.

The proposed algorithm evaluation methodology depends upon a mathematical model that is based on the possibility of predicting statistically the distribution of measured data over some number of channels. Every IR sensor system defines these channels in a natural way, according to the discriminants that it is designed to use. Each pixel in the spatial distribution of an observed scene, the observed signal from each spectrally resolved wavelength band, and each time frame in the temporal sequence of observations constitutes a separate data channel in this sense.

The mathematical model assumes a signal processing logic that divides the decision process for discriminating between targets and background into two steps. The first is the detection phase, which eliminates as many false alarms as possible by means of one or more preliminary target detection algorithms. The second is the declaration phase which generates the final decision as to the presence or absence of a target in a given direction.

A preliminary detection algorithm, used in the first phase, is a linear or nonlinear digital filtering operation followed by thresholding. Tracking algorithms, which distinguish between the resulting target and clutter detections by means of their supposedly different trajectory characteristics observed over time, provide the final, second phase, decision whether or not to declare that a target is present.

For the preliminary detection phase the mathematical model assumes that the statistical distribution of IR radiance over the data channels is adequately approximated by an N-variate Gaussian probability distribution.* There are several arguments to justify this assumption.

*This is a generalization of a similar model proposed in Ref. 2 for spectral discrimination.

First, experimental evidence suggests (Refs. 8 and 9) that for a variety of terrain backgrounds,* although by no means all, in selected spectral channels distributed over a band between 2 μ and 11.4 μ Gaussian distributions fit measurement data remarkably well. This is true for data taken over background regions that comprise as many as two-hundred-thousand pixels.

Second, although, as R. A. Steinberg has pointed out, the mean background radiance can be expected to vary over space and time, the variation is usually gradual except for cases in which glint dominates.** Thus, the assumed N-variate Gaussian distribution can be regarded as a local approximation to the actual N-channel background distribution, valid to the second order in terms of moments of the corresponding density functions.

It is sometimes argued that, although a distribution may be approximately Gaussian out to 2 or 3 σ, acceptable IR sensor system false alarm rates in practice are so low that the tail of the distribution is also significant. This would be true if an attempt were made to meet the false alarm specification with preliminary detection algorithms alone.*** However, in most cases those algorithms are used primarily to thin out the false

---

* Unfortunately, the argument is limited in scope by the fact that similar data for cloud backgrounds does not exist in the literature at the present time.

** In Ref. 18, Steinberg, taking into account photon fluctuations, analyzes the design of optimum filters for thresholding against different spatial variations of a background. His design concepts, as well as other adaptive thresholding techniques, some of which have been implemented in existing IR systems, depend by implication on the assumption that the background variation will be gradual for the most part.

*** Reference 2, in fact, proposes a 12-color spectral detection processing scheme that would do just that if the target and background distributions happen to fit certain models that the authors of the report have generated synthetically and which assume N-variate Gaussian distributions over the 12 channels.

alarms during the preliminary detection phase, and the responsibility for the final target declaration is reserved for tracking algorithms. The burden of satisfying the false alarm rate requirement then rests ultimately on the tracking algorithms.

Perhaps the most important argument for assuming Gaussian distributions, however, is that they furnish a minimum standard of acceptance. That is, a signal processing scheme ought to be regarded as unacceptable if it does not perform well against a Gaussian distributed background. Of course, the converse statement is false; therefore, even if the scheme does meet the standard there may still be cause to reject it, at least for some applications.

In this connection, it should be noted that it is possible to include in an evaluation based on such a minimal acceptance standard the effect of different scenarios which may imply not only a change in the background, e.g., from sky to terrain, but changes in other environmental factors as well. For example, Ref. 4, using calculations obtained from a computer program (5 cm^{-1} LOWTRAN5) for estimating propagation effects, discusses the influence that range and the altitudes of both the target and the sensor platform may have on spectral discriminants. This influence, it is pointed out, would necessarily be reflected in the evaluation of a target detection algorithm, particularly one that relied upon data from multiple spectral channels.

Chapter II of this paper describes in more detail the proposed mathematical model for evaluating discrimination algorithms, as well as some of the model's ramifications when applied to spatial discriminants in particular. The discussion in Chapter II covers the explicit form of the model for both spatial and spectral channels when targets are present or absent. It also indicates how the extensive measurement data presented in Ref. 9 can be used to test the validity of a class of occasionally encountered hypotheses concerning the nature of background spatial distributions for natural scenes.

Chapter III deals with optimum constant false alarm rate (CFAR) detection. It also considers the relationship between the effectiveness of preliminary CFAR detection algorithms and the effectiveness of tracking algorithms used for the final target declaration.

The general optimum CFAR processing rule presented in Chapter III is essentially that found in Ref. 19, which, however, refers to an earlier reference for its derivation. For the sake of completeness an independent derivation of the rule is given in Appendix A.*

Chapter IV analyzes two signal processing techniques that are sometimes encountered in IR processor system designs. One is a method for adaptive thresholding against spatially varying backgrounds; the other is a device to reduce the number of degrees of freedom to be considered in spectral discrimination.

This paper does not include numerical applications to specific cases, except for one or two examples provided to illustrate a point. However, the analysis that is applied to developing the methodology for evaluating discrimination algorithms leads naturally to some conclusions of a general nature which are noted in the text as they occur. These conclusions also appear in Chapter V, along with a summary of the principle ideas introduced in the earlier chapters.

*This might also have been done for a fundamental theorem, introduced in Chapter II, concerning the probability distribution that results from a linear transformation of variables having an N-variate Gaussian distribution. The theorem, however, is reasonably well-known and is heuristically evident.

II. MATHEMATICAL MODELS

A. DATA CHANNELS

An IR sensor with multiple detectors provides data that are separated naturally into discrete channels, each of which is associated with the output signal from one of the detectors. For signal processing purposes, however, it is useful to separate the data into channels that are defined in terms of the discriminants used by the sensor system for distinguishing between target and clutter sources.

Multi-color systems, i.e., systems that rely upon spectral signatures with components in two or more distinct wavelength bands, are the usual examples in which data are treated from this point of view. However, it can be equally useful to regard data as separated into spatial as well as spectral channels, a point of view which this paper will adopt to some advantage, for example, in discussing the effects of linear spatial filtering.

The individual pixels in the background radiance scene mapped by an IR sensor will determine the spatial channels as perceived here. Actually, the number of such channels will generally be limited by an n by n pixel sliding window.* The window defines n^2 spatial channels, one for each pixel contained within it, and is itself defined by whatever spatial filtering algorithms the sensor may use for signal processing.

*The window could just as easily be rectangular. The implicit assumption here that it is square is made for convenience, to simplify to some extent the algebraic treatment of two-dimensional arrays of channels.

It is convenient to require that n be an odd number because the pixel at the center of the array will have a special role in the mathematical model to be proposed here for characterizing spatial discriminants. Specifically, if a target signal occurs in the central channel the data in the full complement of n^2 spatial channels will be regarded as due to the presence of a target. Otherwise, the target will be regarded as absent. It is assumed that the detection algorithm, to the extent that it is based on accurate knowledge of the target and clutter background statistics, is deliberately designed to announce that a detection has occurred if, and only if, the target signal is in the central channel.

This convention implies a desirable, although not necessarily achievable, precision in the location of a target by the IR system. That is, as the array window scans the background a true detection occurs only when the target coincides with the central pixel.

B. NOTATION FOR THE SPATIAL DISCRIMINANT MODEL

In general, data divided among several channels will be treated as a vector each of whose components is the signal strength associated with one of the channels. Unfortunately, the single subscript notation ordinarily used in dealing with a vector $\underset{\sim}{V}$ in terms of its components V_i conflicts with the double subscript matrix notation that is more natural in dealing with the two-dimensional array of signal strengths S_{ij} associated with an n by n array of spatial channels.

Reference 1 (p. 128) handles this problem by providing a so-called stacking transformation that reorders the elements of the array so that they constitute a one-dimensional sequence which can be treated as the components of a vector in the conventional format. Since the transformation is linear and invertible, it is possible to apply standard algebraic manipulations to the vector and change back to the two-dimensional array format whenever it is convenient to do so.

For the purpose of this paper, however, the stacking transformation seems an unnecessary complication that would obscure certain geometric patterns or effects that result, for example, when two or more linear filtering processes are combined. Instead, a two-component vector subscript will be introduced in place of the pair of subscripts ordinarily used to designate an array element. That is, S_{ij} becomes S_k, where the subscript k is regarded as a vector with the components i and j.

In this notation a sum over k will mean a double sum taken independently over all values of i and j. Also, the usual conventions that apply to vectors apply to vector subscripts. Thus, if two vector subscripts are equal it will mean that their corresponding components are equal, and when the vector subscript is 0 it will mean that both subscript components are zero.

It is then possible to represent the linear transformation of an array in the usual manner as a multiplication of a vector by a matrix. That is, a linear transformation from the array with elements S_{ij} to one with elements S'_{ij} will take the form

$$S'_k = \sum_l M_{kl} S_l \quad ,$$

where k and l both represent two component vector subscripts. The matrix with elements M_{kl} then actually has n^4 elements, and the symbol M_{kl} may be understood to have four scalar subscripts.

Sometimes it is necessary to deal with array vectors, or transformations of the type just described, whose algebraic representations depend in some explicit way on their subscripts. When this happens it is usually possible to express such quantities in dyadic form, so that despite the use of vector subscripts it is no more difficult to perform explicit algebraic

manipulations with them than would be the case if their subscripts represented ordinary scalar integers.

In order to emphasize its key role, the central channel in an n by n array will be designated by the zero vector subscript, which is equivalent to two zero scalars. Then, with scalar subscripts ordered in the standard manner, with the conventional reference to an array element's position by row and column, negative subscripts will be used to designate elements to the left of or above the center. That is, for an element S_{ij}, i and j will both range over the integers from $\frac{1-n}{2}$ to $\frac{n-1}{2}$. For example, in the case n=3, that is, for a 3-by-3 or 9-element array, the array would have the form

$$\begin{pmatrix} S_{-1-1}, & S_{-10}, & S_{-11} \\ S_{0-1}, & S_{00}, & S_{01} \\ S_{1-1}, & S_{10}, & S_{11} \end{pmatrix} .$$

C. PROBABILITY DISTRIBUTIONS

One way to interpret the problem of detecting the presence of a target against a clutter background is to regard it as the problem of estimating the probability that the target is present, given the information acquired from the data provided by IR measurements. On the basis of this concept Ref. 2 has introduced a minimum error criterion* for multi-color systems to distinguish between targets and clutter by means of their spectral characteristics.**

---

*A minimum error criterion in this context is one that classifies each signal as due either to the target or to clutter alone with the smallest possible probability of an erroneous classification. Cf. Ref. 3, p. 269ff.

**See also the discussion in Ref. 4, Appendix B.

In many applications, especially those involving point targets, however, the false alarm rate is a major concern. It is a primary objective of the present paper to formulate a method for evaluating target detection algorithms when this is, in fact, the case. Accordingly, a related but slightly different approach will be taken here. The concepts underlying this approach can be summarized as follows.

For an N channel system each measurement set produces an N component vector which may be thought of as representing a point in N dimensional data space. The set of all such data points that might be produced by clutter in the absence of a target has, at least conceptually, an N-variate joint probability distribution defined by a probability density function $P_C(\underset{\sim}{J})$, where $\underset{\sim}{J}$ is a vector having components $J_1, \ldots, J_N$ that may, individually, range over all positive and negative real values. Similarly, there is another such probability distribution, and a corresponding density function $P_T(\underset{\sim}{J})$ that is associated with the presence of a target source.

Suppose that there is an algorithm whose purpose is to decide whether a given measurement set, i.e., data point, was produced in the presence or absence of a target. The algorithm then has the effect of separating all of data space into two complementary regions.

One of the regions R will consist of all points designated by the algorithm as due to a target. The other will consist of all points designated as due to clutter in the absence of a target.

The probability of false alarm (PFA) for any measurement set is then equal to the integral of $P_C(\underset{\sim}{J})$ over R; i.e.,

$$PFA = \int_R P_C(\underset{\sim}{J})\, dJ_1 \ldots d J_N \quad . \tag{1}$$

Also, the probability of detecting a target (PTD) by means of the algorithm applied to a single measurement set is equal to the integral of $P_T(\underset{\sim}{J})$ over R; i.e.,

$$PTD = \int_R P_T(J)\, dJ_1 \ldots d J_N \quad . \tag{2}$$

Throughout this paper it will be assumed that $P_C(\underset{\sim}{J})$ and $P_T(\underset{\sim}{J})$ are both N-variate Gaussian probability density functions.* That is, each will have the form

$$P(\underset{\sim}{J}) = (2\pi)^{-\frac{N}{2}} \left|\underset{\approx}{M}\right|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\,(\underset{\sim}{J}-\bar{\underset{\sim}{J}})^t\, \underset{\approx}{M}^{-1}\, (\underset{\sim}{J}-\bar{\underset{\sim}{J}})\right] , \tag{3}$$

where $\underset{\approx}{M}$ is the covariance matrix of the particular distribution, $|\underset{\approx}{M}|$ is the determinant and $\underset{\approx}{M}^{-1}$ the inverse of $\underset{\approx}{M}$, $\bar{\underset{\sim}{J}}$ is the mean vector of the distribution, and the superscript t denotes the transpose. In (3) ordinary matrix multiplication is implied, so that a vector without a superscript is to be regarded as a column vector while one that has the superscript t is to be regarded as a row vector.

The covariance matrix and mean vector are the parameters that specify a particular N-variate Gaussian distribution. Therefore, when specific reference is made to P_C or to P_T in the expression for the density given by (3) $\underset{\approx}{M}$ and $\bar{\underset{\sim}{J}}$ will bear the appropriate subscript, C or T.

There are several arguments in favor of assuming Gaussian probability distributions for the measured signal strengths. Since a major objective of this paper is to devise a method for testing clutter rejection algorithms analytically, the principal argument is that the very least one might expect from such an algorithm would be satisfactory performance when the target and clutter signal strengths are Gaussian distributed.

*For properties of N-variate Gaussian probability distributions see, for example, Ref. 5, Chs. 21-24 or Ref. 6.

There are certainly a number of environments for which existing empirical data suggest that the assumption of Gaussian statistics may be surprisingly accurate. Examples occur in the data considered by Ref. 7--notably, that taken from Ref. 8 and particularly that from Ref. 9, which will be discussed in the next section.

D. EXPERIMENTAL DATA AND EXISTING STATISTICAL MODELS

The amount of data collected through IR measurement over the years is voluminous. Measurement programs for this purpose have covered a variety of targets and clutter backgrounds in virtually all spectral bands of practical interest. Reference 7 contains an in-depth survey of the most important experimental results derived from such programs and also provides a detailed analysis of how the data may be affected by environmental factors.

Unfortunately, of the many sources available in the literature, only Ref. 9 offers data processed in a form that is directly applicable to the mathematical models used in this paper. What is needed particularly are means and covariance matrices, the elements of which depend upon the standard deviation for each channel and the correlation coefficients between all pairs of channels. It is unfortunate that data for cloud backgrounds have not been published in a similar form.

Reference 9 provides all of these parameters for several spectral channels* generated by a number of different terrain backgrounds, each observed during four time periods--predawn, noon, sunset, and midnight. The observations were made from an airborne platform at 90 deg and 35 deg depression angles with instantaneous fields of view (IFOV) ranging from 2 to 5 mrad at altitudes from 1,000 ft to 1,750 ft. However, only terrain backgrounds were measured; no examples of sky, clouds or ocean are included in the collection.

*Of course, the measurements were made in a particular set of fixed wavelength bands. However, Ref. 9 recommends a method of interpolating the measured data to derive equivalent approximate data for other choices of spectral decomposition.

Aside from the first and second moment statistical parameters, for each case Ref. 9 also presents the data in several other forms. These include: (1) a histogram for each spectral band, along with an overlay of the Gaussian probability density curve defined by the mean and standard deviation associated with the histogram, (2) area diagrams showing the size and orientation of all subregions with radiance above a 2σ and above a 3σ threshold, (3) both the cross-track and in-track power spectral densities (sometimes called the Wiener spectra) for the measured region. Figures 1-6, taken from Ref. 9, are examples of all three graphic forms of data.

In many of the cases presented in Ref. 9 the Gaussian density curve fits the corresponding histogram with remarkable accuracy out to the 2, 3, and sometimes even the 4σ level. This is especially true for midnight scenes that are natural in origin, such as a conifer forest or a desert, as distinguished from land or cities. Figures 1 and 2 show that the fit is fairly good for a conifer forest even at noon.

Other histograms are multi-modal and skewed. However, for many of these, in the accompanying area diagrams that display the thresholded subregions of maximum radiance, the high-temperature zones appear to be relatively isolated and confined to one or two small areas in the overall background.* When this is the case it seems likely that the lesser modes appearing in the histogram tail would not be present if the scene were broken up into smaller regions and a separate histogram of the radiance distribution were constructed for each of the newly formed regions.

For other cases, e.g., the city of Baltimore, Maryland and Fort A.P. Hill, Virginia, to name the most extreme examples, the multi-modal character of the histogram is evidently not the result of isolated effects in the background. Instead, the high-temperature zones are distributed throughout the scene, and therefore it must be concluded that a Gaussian distribution will not adequately represent these data.

*Cf. Figs. 3 and 4.

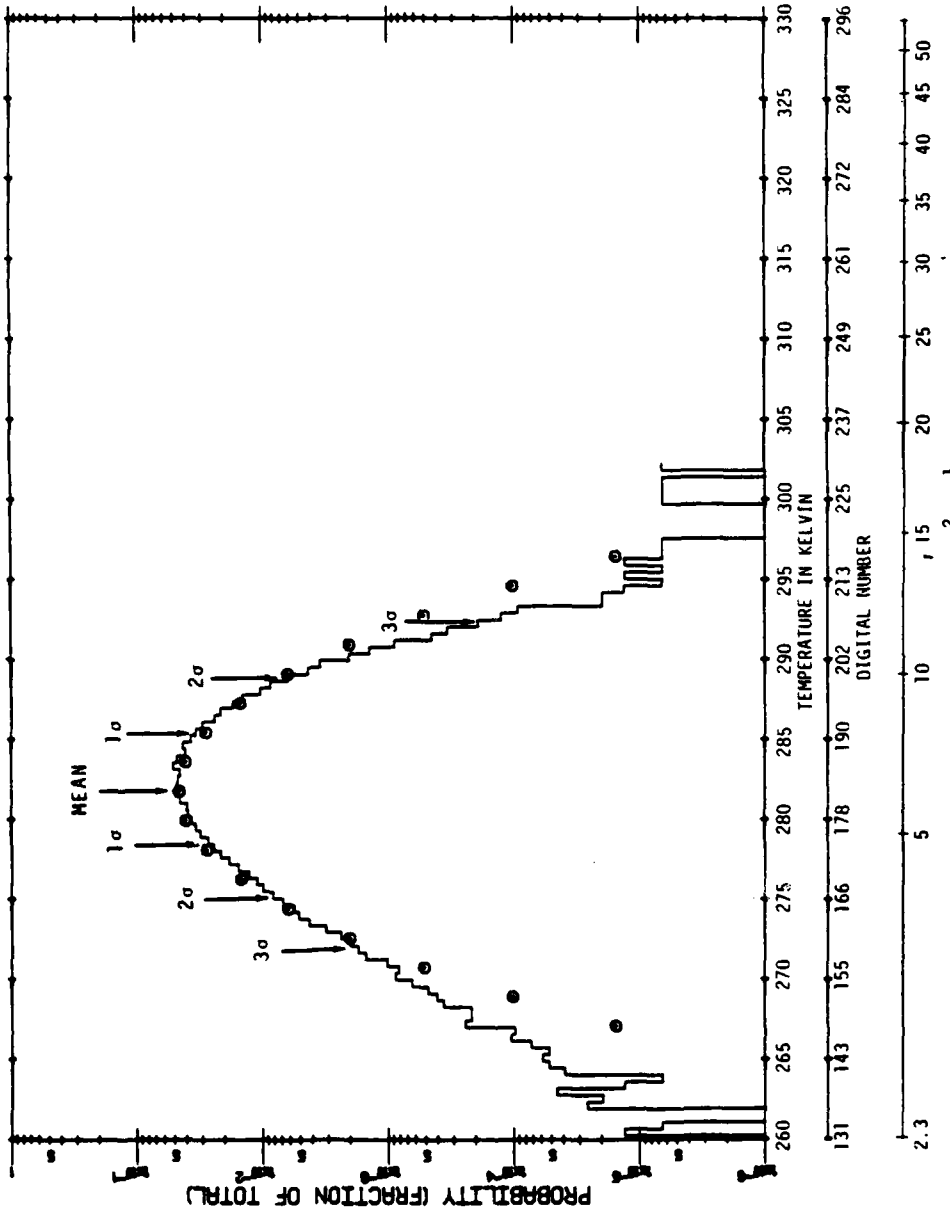


FIGURE 1. Histogram of Michigan winter scene - Noon (Angle: 90 deg.)

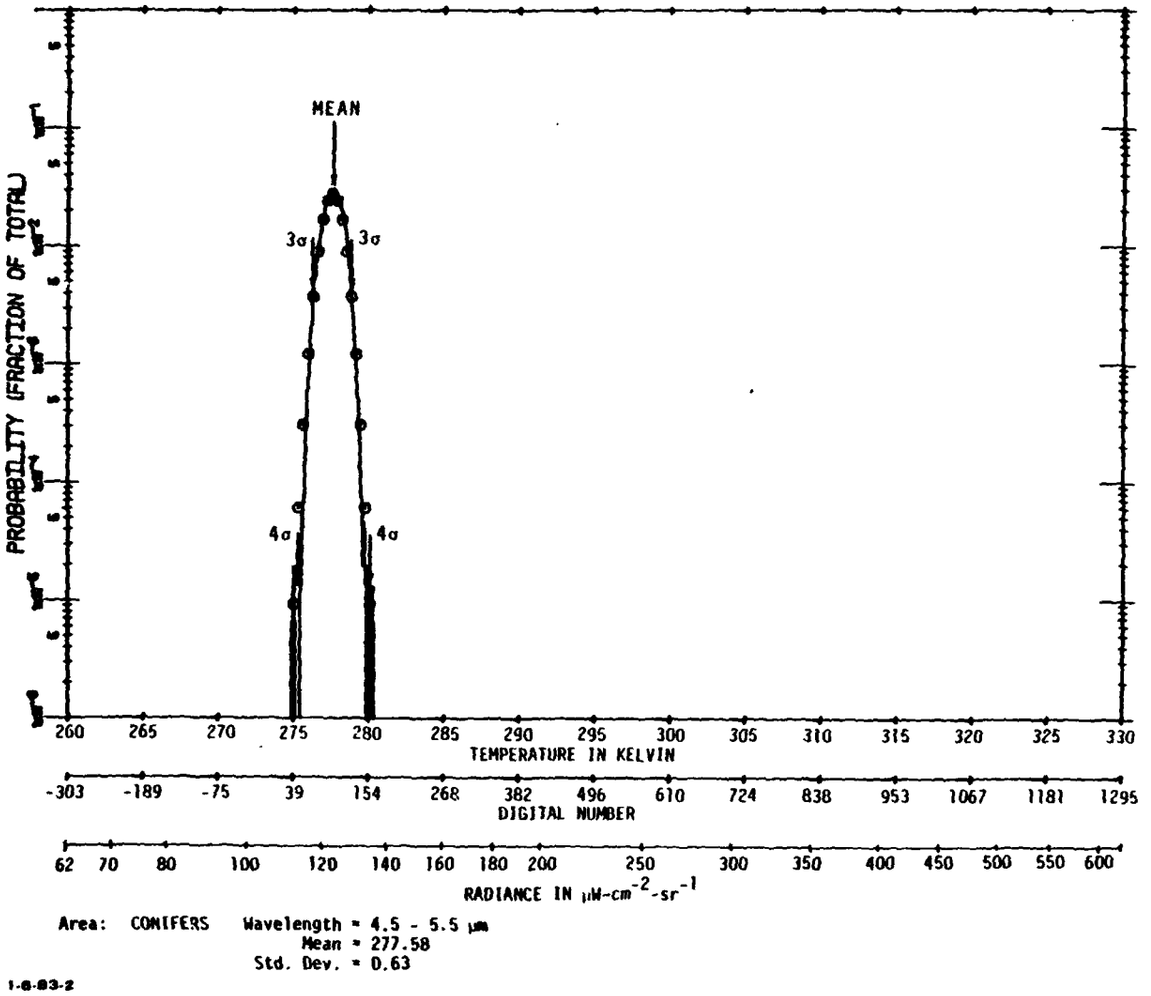
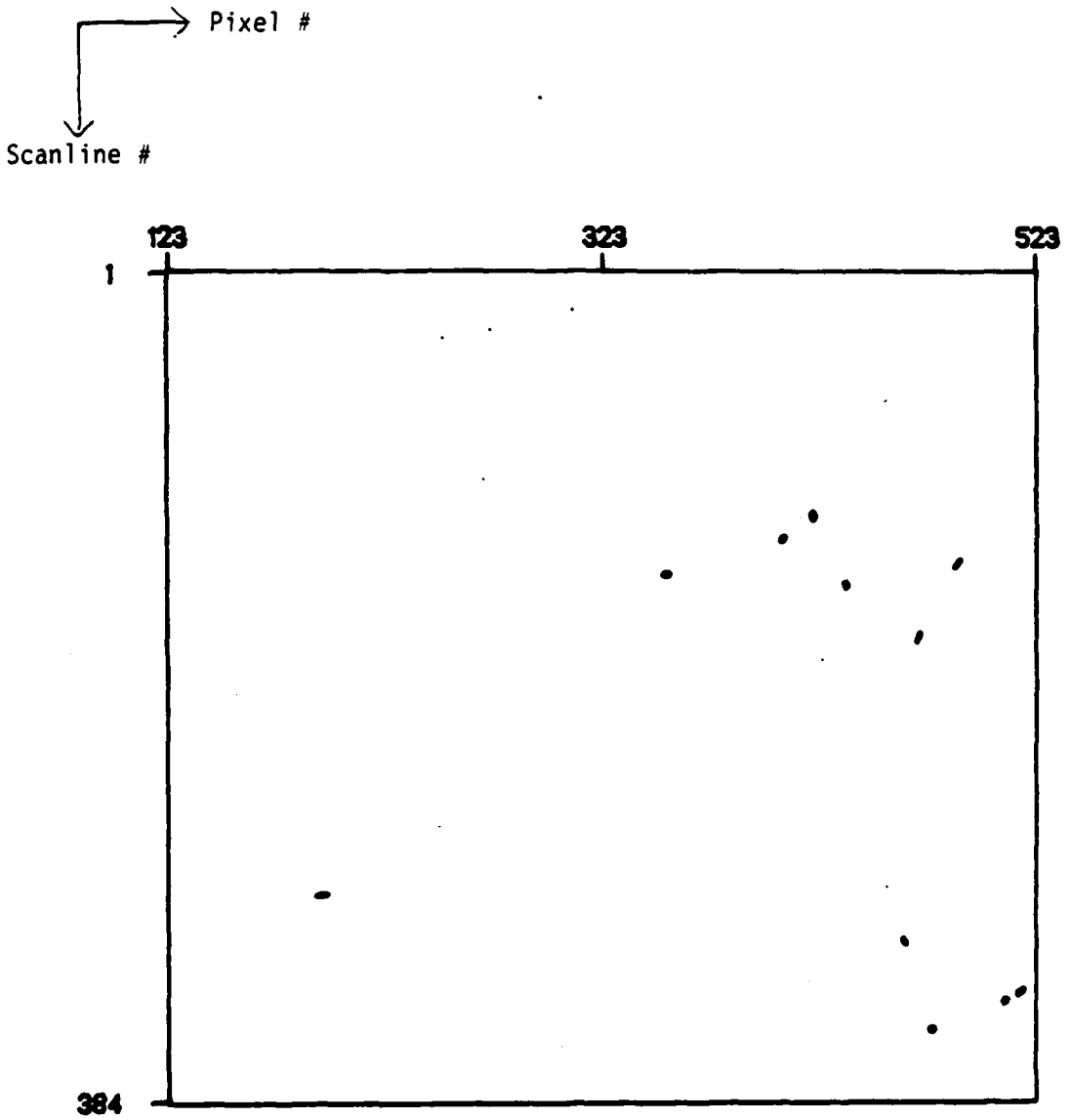


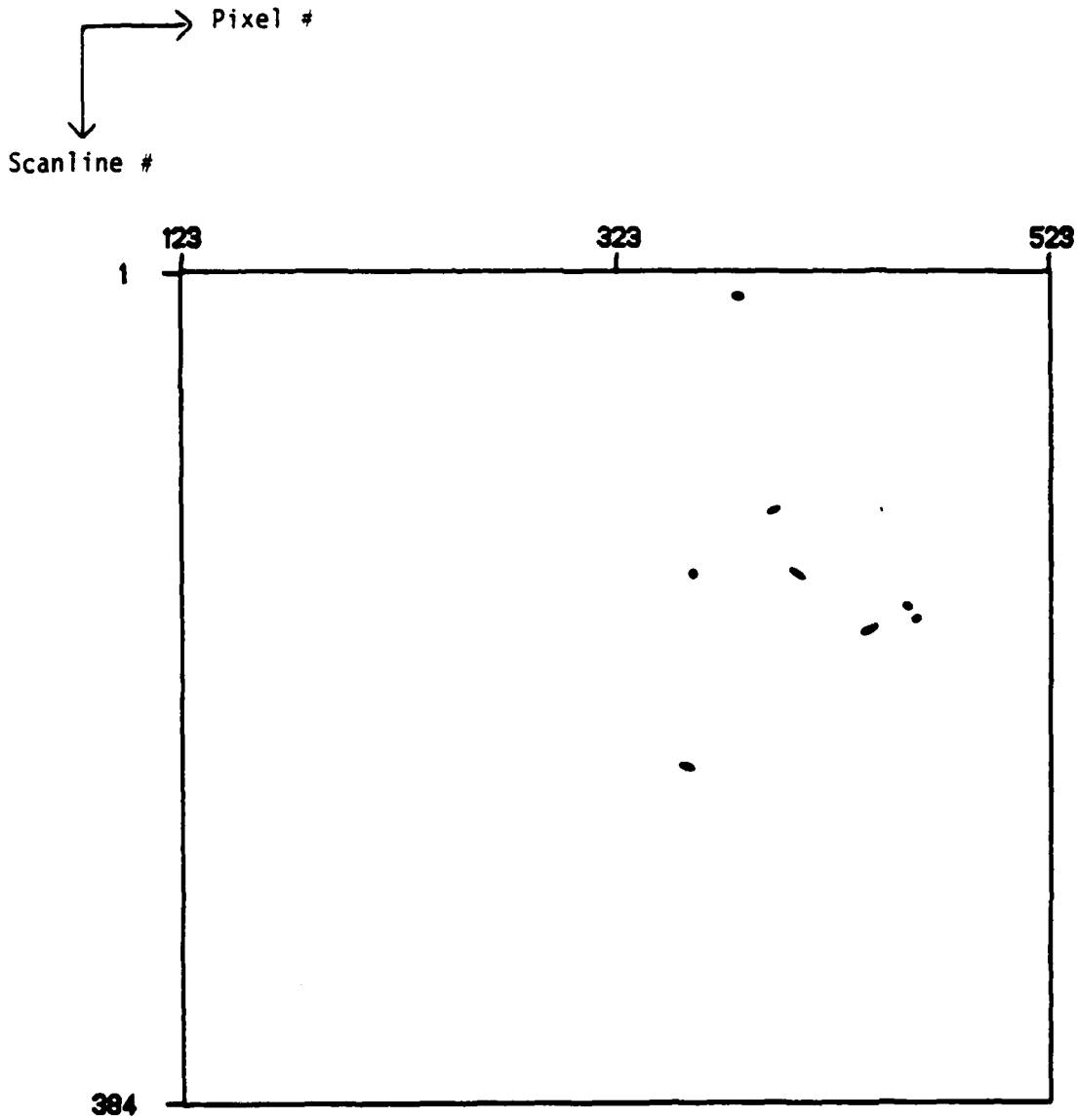
FIGURE 2. Histogram of Michigan winter scene - Noon (Angle: 90 deg.)



Area: CONIFERS
Temperature Threshold
= Ave. + 2.00 σ
Wavelength = 3.5 - 3.9 μm

1-6-83-3

FIGURE 3. Equivalent elliptical areas for Michigan winter scene - noon



Area: CONIFERS

Temperature Threshold

= Ave. + 2.79 σ

Wavelength = 4.5 - 5.5 μm

1-6-83-4

FIGURE 4. Equivalent elliptical areas for Michigan winter scene - noon

Reference 2 details the construction of a set of quasi-synthetic statistical models for the spectral distribution of radiance due to a variety of potential targets and clutter sources. These models were developed by means of analysis based on physical principles and what appear to be reasonable assumptions combined with empirical data gathered from a number of different references, including Ref. 9.

An important application of this work is embodied in a computer program called PALANTIR, which Ref. 2 also describes in some detail. From a given set of narrow band spectral channels PALANTIR chooses a prescribed number of channels, picking those that will provide the least error when used in connection with a minimum error algorithm for discriminating between targets and clutter. The basis for this choice is a test which depends upon the means and covariance matrices associated with the statistical models.

In an attempt to construct a theoretical model for spatial channels, Ref. 10 postulates statistical homogeneity for terrain backgrounds, citing as evidence for this assumption IR measurements taken by the Lincoln Laboratory at 20 natural settings in New England. Statistical homogeneity in this case means that for the radiance distribution spatially the cross-correlation between any two pixels depends only on the amount of their separation and not on the position of either in the scene.

A further assumption of Ref. 10, for which the same evidence is cited, is that the cross-correlation is an exponential function of the separation. That is, for radius vectors $\underset{\sim}{r}$ and $\underset{\sim}{r}'$ that determine two points in the plane of the radiance distribution it is assumed that the cross correlatcn $K(\underset{\sim}{r}, \underset{\sim}{r}')$ between the radiance values at the two points has the form

$$K_{x,y}(\underset{\sim}{r}, \underset{\sim}{r}') = \exp\left(-\frac{|x-x'|}{L_x} - \frac{|y-y'|}{L_y}\right), \qquad (4)$$

where L_x and L_y are correlation distances in the x and y directions and (x,y) and (x', y') are the respective components of $\underset{\sim}{r}$ and $\underset{\sim}{r}'$.

The model recommended by Ref. 10 for spatial correlation is general enough to provide for anisotropic behavior; however, its functional form obviously depends upon the choice of the coordinate system. If the spatial distribution were also assumed to be isotropic the correlation function would be independent of the coordinate system. If it were also exponential it would have the form

$$K(\underset{\sim}{r}, \underset{\sim}{r}') = \exp\left(-\frac{|\underset{\sim}{r}-\underset{\sim}{r}'|}{L}\right) \quad , \tag{5}$$

which is completely determined by a single correlation distance L.

It is interesting to note that Ref. 10 assumes the anisotropic form (4) for the cross-correlation because the cited supporting data were measured at a depression angle of 20 deg. The argument is that one might expect a scale change from in-track to cross-track linear distance measurements relative to a statistically homogeneous two-dimensional distribution because of the distortion created in the cross-track direction by the depression angle.

However, if the appropriate form of the cross-correlation to account for this distortion were indeed (4) as assumed, then for a 90-deg depression angle the cross-correlation would be given by

$$K_{x,y}(\underset{\sim}{r}, \underset{\sim}{r}') = \exp\left[-\left(\frac{|x-x'| + |y-y'|}{L}\right)\right] \quad ,$$

which is still anisotropic despite the single correlation distance parameter L.

The two-dimensional Fourier transform of either correlation function, given by (4) or (5), is the corresponding power spectral density or Wiener spectrum, $W_{x,y}(\underset{\sim}{k})$ or $W(\underset{\sim}{k})$, in terms of a vector wave number $\underset{\sim}{k}$. The two densities are given by*

$$W_{x,y}(\underset{\sim}{k}) = \frac{4\, L_x L_y}{(1+k_x^2\, L_x^2)(1+k_y^2\, L_y^2)} \quad , \tag{6}$$

$$W(\underset{\sim}{k}) = \frac{2\pi L^2}{(1+k^2\, L^2)^{3/2}} \quad , \tag{7}$$

where k_x and k_y are the Cartesian components and k is the magnitude of the wave number vector.

In principle, (6) or (7) might be used to check whether either of the corresponding correlation functions is a good model for a given background when the data obtained from measurements of the background include linear components of the Wiener spectra in at least two different directions. In fact, Ref. 9 does provide data in this form for every case considered and for correlations in both cross-track and in-track directions relative to the scanning motion of the sensor. However, there is no reason to believe that the track direction coincides with either the x or y direction, both of which may be at least partially determined by the physical properties of the background distribution rather than by the motion of the sensor.

Nevertheless, a comparison of the cross-track power spectrum curve with the in-track curve affords at least a preliminary check on the possibility that the radiance distribution is isotropic, i.e., by observing whether the curves are nearly the same. Examples of distributions that may be isotropic do exist in the Ref. 9 data, e.g., for a conifer forest background observed at a 90-deg depression angle at noon.

*See Appendix B.

Figures 5 and 6, taken from Ref. 9, contain Wiener spectra for this case. Curves for the 3.5-3.9 μ and 4.5-5.5 μ bands roughly approximating the in-track spectra depicted in Fig. 6 are shown as dashed lines in Fig. 5 to illustrate the point.

However, an examination of the conifer forest power spectrum curves presented in Ref. 9 fails to disclose any that might correspond to the functional behavior indicated by (7). In every case the spectral density either decreases too rapidly or too slowly with increasing wave number.

It is possible that by changing exponents in the denominator of (7), e.g., replacing the exponent $\frac{3}{2}$ with 2 or with $\frac{5}{4}$ a better fit to the experimental Wiener spectra might be obtained; Appendix B shows how to calculate the corresponding correlation functions explicitly. Some numerical experimenting with new exponents indicates for the conifer forest data, however, that although changing exponents in (7) can improve the fit somewhat, at best it can only be made close at two points on a given curve.

E. SIMPLE MODELS FOR TARGET STATISTICS

In principle, it is possible for a sensor to estimate the mean and the covariance matrix elements for clutter statistics by making sample measurements of the background before a target arrives and updating these estimates periodically. But it is even conceptually difficult to imagine how this information might be obtained for targets in general. The possibility of using a predetermined catalogue of signatures for this purpose seems limited because of the many variations in range, aspect, altitude, velocity, and position of a target relative to the sun.

Fortunately, in the case of point targets it is usually reasonable to assume that covariance matrix elements will be dominated by background statistics. To the extent that this is true, for discrimination purposes it is only necessary to anticipate the mean values associated with the data channels that define a target's signature.

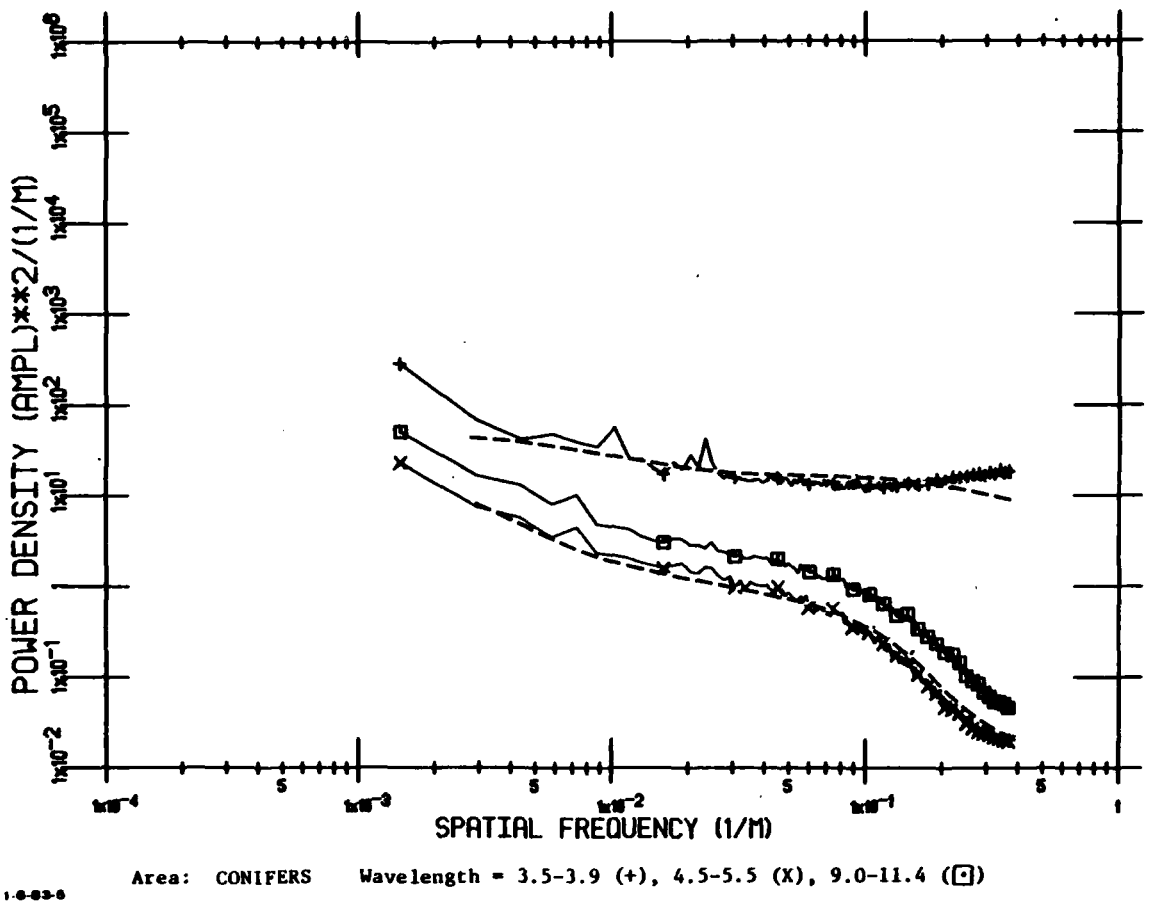
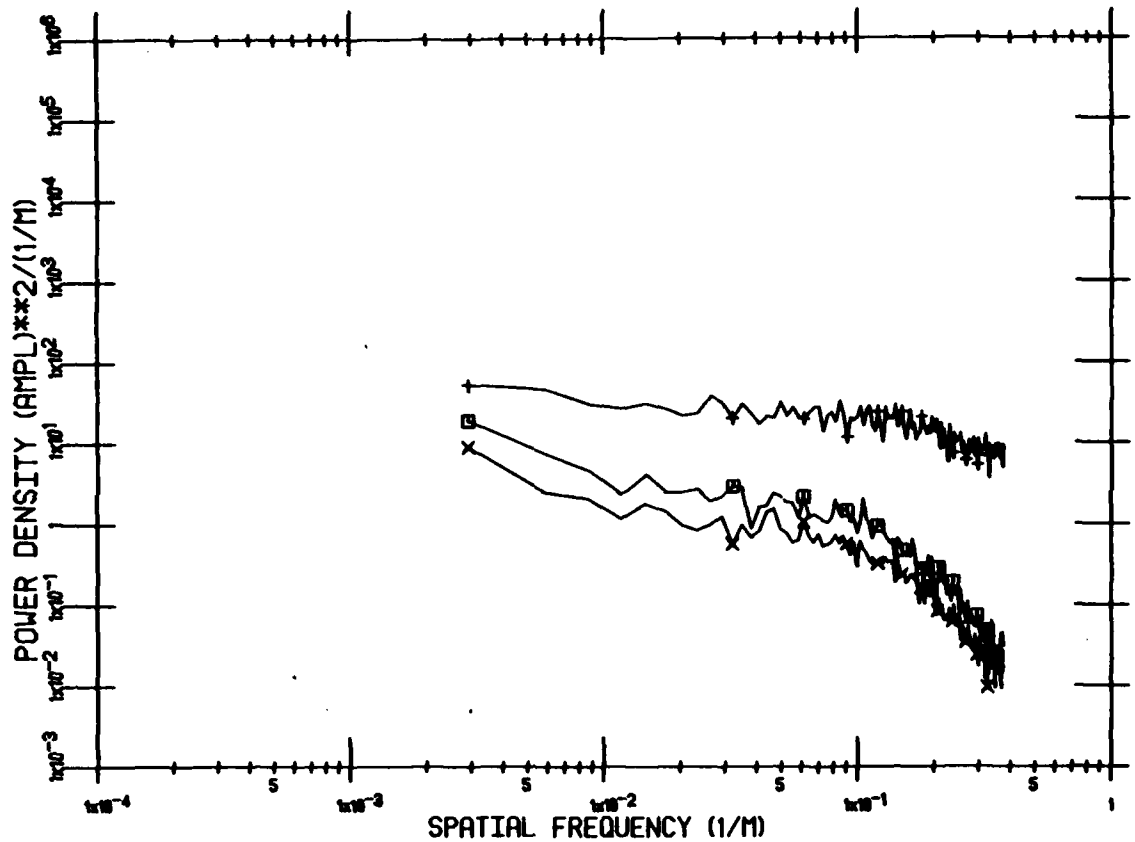


FIGURE 5. Power spectra - Michigan winter scene: Noon - (Angle: 90 deg.) - Crosstrack



Area: CONIFERS Wavelength = 3.5-3.9 (+), 4.5-5.5 (X), 9.0-11.4 (⊡)

1-e-83-e

FIGURE 6. Power spectra - Michigan winter scene: Noon - (Angle: 90 deg.) - Intrack

As remarked in Ref. 7, for spectral channels the observed radiance is essentially the sum of two parts: the first is due to the background except for the area occulted by the target; the second is the difference between the radiance due to the target and the background radiance that would result from the occulted area if it were not obscured.

Included in the radiance there should be a part due to atmospheric emissions along the propagation path. However, it will be assumed here that this contributes a negligible amount to statistical fluctuations about the mean.

In the case of a point target, which, by definition, occupies only a small part of the sensor's footprint, the mean radiance observed is that obtained from a calculation of the type suggested in Ref. 7. The calculation is equivalent to a weighted average $\bar{J}_T$, given by

$$\bar{J}_T = W_C \bar{J}_C + W_T \hat{J}_T \quad , \tag{8}$$

where $\hat{J}_T$ is the radiance supplied by the target, $\bar{J}_C$ is the mean clutter radiance, and the two coefficients W_T and W_C are fractions of the total footprint area within and without the occulted area.

According to a well-known theorem,* if the components of an N-dimensional vector $\underset{\sim}{J}$ have an N-variate Gaussian joint probability distribution with the mean vector $\underset{\sim}{\bar{J}}$ and the covariance matrix $\underset{\approx}{M}$, then the components of the M-dimensional vector $\underset{\sim}{Y}$ resulting from the linear transformation obtained when $\underset{\sim}{J}$ is multiplied on the left by an M by N matrix $\underset{\approx}{T}$ (i.e., $\underset{\sim}{Y} = \underset{\approx}{T}\,\underset{\sim}{J}$) will have an M-variate Gaussian joint probability distribution with the mean vector $\underset{\sim}{\bar{Y}}$ given by

$$\underset{\sim}{\bar{Y}} = \underset{\approx}{T}\,\underset{\sim}{\bar{J}} \tag{9}$$

and the covariance matrix

$$\underset{\approx}{M}_y = \underset{\approx}{T}\,\underset{\approx}{M}\,\underset{\approx}{T}^t \quad .$$

*Ref. 6, p. 86.

Given the present assumptions, the background radiance J_C and the target radiance $\hat{J}_T$ may be regarded as having a joint bivariate Gaussian probability distribution for which the mean vector has the components $\bar{J}_C$ and the mean of $\hat{J}_T$ and the covariance matrix has the form

$$M = \begin{pmatrix} \sigma_C^2 , & 0 \\ 0 , & 0 \end{pmatrix} \quad . \tag{10}$$

In (10) it is, of course, implicit that the target and background radiances $\hat{J}_T$ and J_C are uncorrelated. Also, (10) represents a limiting case in which the standard deviation of $\hat{J}_T$ is vanishingly small, so that there are no fluctuations of $\hat{J}_T$ about its mean.

In accordance with (8) and the first equation in (9) the vector with the components W_C and W_T corresponds to a 1 by 2 transformation matrix $\underset{\approx}{T}$. Then, according to the cited theorem and (10), the combined target and background radiance will have a univariate Gaussian probability distribution with a mean $\bar{J}_T$ given by (8) and a variance σ_T^2 given by

$$\sigma_T^2 = W_C^2 \, \sigma_C^2 \quad . \tag{11}$$

In the case of multiple spectral channels J_C, $\hat{J}_T$ and, inferentially, $\bar{J}_C$ would all be replaced by vectors in (8). Then (11) would be unchanged in form except that σ_C^2 would be replaced by a matrix; σ_C^2, in fact, would be replaced by the background covariance matrix associated with the multiple channels. That is, for N spectral channels

$$\begin{aligned} \underset{\sim}{\bar{J}}_T &= W_C \, \underset{\sim}{\bar{J}}_C + W_T \, \underset{\sim}{\hat{J}}_T \quad , \\ \underset{\approx}{M}_T &= W_C^2 \, \underset{\approx}{M}_C \quad , \end{aligned} \tag{12}$$

where $\hat{J}_T$, $\bar{J}_C$ and $\bar{J}_T$ are N component vectors, $\underset{\approx}{M}_C$ is the N by N background covariance matrix, and $\underset{\approx}{M}_T$ is an N by N matrix that may be regarded as the effective target covariance matrix. Acccrding to (12) the target and background covariance matrices are proportional.

A similar model can be devised for spatial channels that form an N by N pixel array. Using C to denote a clutter source, T to denote a target source, and the case N = 3 for illustration, when the target is absent the pixel array will have the form

$$\begin{pmatrix} C, C, C \\ C, C, C \\ C, C, C \end{pmatrix}$$

and when the target is present the form

$$\begin{pmatrix} C, C, C \\ C, T, C \\ C, C, C \end{pmatrix} .$$

The case of a target source in the array but not at the center will be regarded as a case in which the target is absent. It will be assumed that the probability that a target will be anywhere within the array at any given time is small. Thus, the cases for which it is present but not at the center may be neglected as consisting of an insignificant number of events in comparison with the number of events for which it is absent altogether. That is, such events will have a negligible effect on the clutter probability distribution.

It will also be assumed that the target source is uncorrelated with any background clutter source. Consider, for example, the ideal case in which the target exactly occupies the central

pixel so that the clutter is completely occulted. When this happens the covariance matrix $\underset{\approx}{M}_T$ associated with presence of a target must be such that the cross-correlation between the central pixel and every other pixel in the array is zero.

Then

$$\underset{\approx}{M}_T = \underset{\approx}{M}_C - \Delta\underset{\approx}{M} \quad , \tag{13}$$

where $\underset{\approx}{M}_C$ has the elements M_{ij} and $\Delta\underset{\approx}{M}$ has the elements

$$\Delta_{ij} = M_{jo}\,\delta_{io} + M_{io}\,\delta_{jo} - M_{oo}\,\delta_{io}\,\delta_{jo} - \sigma_T^2\,\delta_{io}\,\delta_{jo} \; . \tag{14}$$

In (14) the notation described earlier, according to which the subscripts are all two-component vectors, is to be assumed. The quantities δ_{ij} are Kronecker deltas, which vanish except when the vectors i and j are identical.

It is easily verified that (13) and (14) define a target covariance matrix with the appropriate properties. That is, when either i or j, but not both, is the zero vector the corresponding pixel is being correlated with the center of the array, and, as it should, the corresponding matrix element of $\underset{\approx}{M}_T$ vanishes. When neither i nor j is the zero vector neither of the pixels is at the array's center, and, as it should be, the corresponding element of $\underset{\approx}{M}_T$ is identical with that of $\underset{\approx}{M}_C$. Finally, when both i and j are the zero vector the pixel is at the array's center, and the corresponding element of $\underset{\approx}{M}_T$ is σ_T^2 which, as it should be, is the variance of the target source.

For the case in which the target source occupies only part of the central pixel the terms in (14) will be weighted as in (12). Analysis similar to that used to derive the weight factors

for spectral channels will provide the appropriate weight factors for spatial channels.*

One other concern in modeling the signal produced by an IR sensor should be mentioned. Most systems provide contrast rather than absolute measurements of the radiance distribution in a scene.

Since the contrast is approximately the difference between the radiance values observed at two successive pixels along a scanline in the scene, its measurement is equivalent to applying a linear filter (high pass) to the spatial channels. The effect of linear spatial filtering in general is discussed in Section B of Chapter III.

*If the region occupied by the target source is larger than the central pixel, Δ_{ij} should contain additional terms with factors of the form δ_{i1}, Mil, and weights given by the components of a vector associated with the rows and columns twice removed from the central pixel.

III. CFAR TARGET DETECTION ALGORITHMS

A. OPTIMUM SEGMENTATION

The term "segmentation" is used in image processing literature to denote the process of separating different classes of objects in a scene. Generally, this is done with the purpose of minimizing the probability that there will be an error in the classification. However, for the applications of interest to this paper, in separating targets from clutter it is more important to set a bound on the probability of false alarm. This is equivalent to prescribing a constant false alarm rate (CFAR), which is a goal common to many IR systems.

Given the CFAR condition, the problem of optimizing the segmentation may be restated as follows. Among all possible rules for detecting the presence of a target with a given false alarm probability, find the rule for which the probability of the detection is a maximum.

Appendix A derives the general solution of this problem in terms of the joint probability density $P_C(\underset{\sim}{J})$ for the distribution of radiance values over the available data channels in the absence of a target and the corresponding joint probability density $P_T(\underset{\sim}{J})$ when the target is present. As in Chapter II, $\underset{\sim}{J}$ is an N-dimensional vector each of whose components is the radiance value in one of the channels.

The solution is to declare that a target is present when the measured components of $\underset{\sim}{J}$ are such that $\underset{\sim}{J}$ defines a point in a certain N-dimensional region R_T. The boundary for this region is a hypersurface that is determined by the equation

$$\log P_T(\underset{\sim}{J}) - \log P_C(\underset{\sim}{J}) = \text{constant}. \qquad (8)$$

The constant in (8) is determined by the condition

$$\int_{R_T} P_C\ (\underset{\sim}{J})\ dJ,\ldots dJ_N = \phi \quad , \tag{9}$$

which is equivalent to the CFAR requirement that the probability of a false alarm be equal to ϕ.

For the case of N-variate Gaussian probability densities $P_T\ (\underset{\sim}{J})$ and $P_C\ (\underset{\sim}{J})$ with mean vector $\underset{\sim}{J}_T$ and $\underset{\sim}{J}_C$ and covariance matrix $\underset{\approx}{M}_T$ and $\underset{\approx}{M}_C$, respectively, (8) and (9) reduce to

$$Q(\underset{\sim}{J}) \equiv (\underset{\sim}{J}-\underset{\sim}{J}_T)^t\ \underset{\approx}{M}_T^{-1}\ (\underset{\sim}{J}-\underset{\sim}{J}_T) - (\underset{\sim}{J}-\underset{\sim}{J}_C)^t\ \underset{\approx}{M}_C^{-1}\ (\underset{\sim}{J}-\underset{\sim}{J}_C) = \gamma \ , \tag{10}$$

where γ is a positive constant determined by

$$\frac{1}{\sqrt{(2\pi)^N\ |\underset{\approx}{M}_C|}} \int_{R(\gamma)} \exp\left[-\tfrac{1}{2}\ (\underset{\sim}{J}-\underset{\sim}{J}_C)^t\ \underset{\approx}{M}_C^{-1}\ (\underset{\sim}{J}-\underset{\sim}{J}_C)\right] dJ,\ldots dJ_N = \phi. \tag{11}$$

In (10) $Q(\underset{\sim}{J})$ is obviously a quadratic function of the N components of $\underset{\sim}{J}$, so that the hypersurface defined by (10) is a quadric surface (e.g., a conic section in the case N=2). In (11) the region of integration $R(\gamma)$ consists of all points $\underset{\sim}{J}$ for which

$$Q(\underset{\sim}{J}) < \gamma \quad . \tag{12}$$

The reason why $R(\gamma)$ is defined by (12) instead of by

$$Q(\underset{\sim}{J}) > \gamma$$

is that $\underset{\sim}{J} = \underset{\sim}{J}_T$ satisfies (12) (this follows from the fact that $\underset{\approx}{M}_C$, and therefore $\underset{\approx}{M}_C^{-1}$, must be positive definite); therefore, the mean vector for the probability distribution when a target is present corresponds to a point in the region defined by (12). For a practical case, in which the probability of a target detection, given by

$$PTD = \frac{1}{\sqrt{(2\pi)^N \, |\underset{\approx}{M}_T|}} \int_{R(\gamma)} \exp\left[-\frac{1}{2}\,(\underset{\sim}{J}-\bar{\underset{\sim}{J}}_T)^t \, \underset{\approx}{M}_T^{-1} \, (\underset{\sim}{J}-\bar{\underset{\sim}{J}}_T)\right] dJ, \ldots dJ_N, \quad (13)$$

is large enough to be of any use, the mean vector $\bar{\underset{\sim}{J}}_T$ would have to represent a point in the region $R(\gamma)$ of integration in (13).

The use of (12), subject to the condition (11), as a test to determine whether the presence of a target should be declared is a somewhat less formidable problem numerically when the covariance matrices $\underset{\approx}{M}_T$ and $\underset{\approx}{M}_C$ are both diagonal. This will be true only if all N data channels are mutually independent in the statistical sense.

If the covariance matrices are not diagonal there exists a linear transformation of the vector $\underset{\sim}{J}$ to a vector $\underset{\sim}{J}'$ such that the probability densities $P_C(\underset{\sim}{J}')$ and $P_T(\underset{\sim}{J}')$ will both have covariance matrices that are diagonal. This follows from the theorem used to derive the expressions (9) in Chapter II and the well-known fact that there is always a linear transformation that can diagonalize any two symmetric matrices simultaneously as long as one of them is positive definite.* Appendix C describes the process of finding the required transformation and carries it out in detail for the point target case in connection with spatial channels.

*Cf. Ref. 11, pp. 37-41.

When $\underset{\sim}{M}_T$ and $\underset{\sim}{M}_C$ are both diagonal, (10), which represents a hypersurface, has the form

$$\sum_{n=1}^{N}\left(\frac{1}{\sigma_{Tn}^2}-\frac{1}{\sigma_{Cn}^2}\right)J_n^2-2\sum_{n=1}^{N}\left(\frac{\bar{J}_{Tn}}{\sigma_{Tn}^2}-\frac{\bar{J}_{Cn}}{\sigma_{Cn}^2}\right)J_n+\sum_{n=1}^{N}\left(\frac{\bar{J}_{Tn}^2}{\sigma_{Tn}^2}+\frac{\bar{J}_{Cn}^2}{\sigma_{Cn}^2}\right)=\gamma, \quad (14)$$

By setting all of the J_n in (14) equal to zero except for two values, ν and μ, it is possible to obtain the two-dimensional cross-section of the region bounded by the hypersurface (which is, itself, an N-1 dimensional manifold) in the ν, μ plane. This cross-section will be the region bounded by the curve

$$K_\nu J_\nu^2+K_\mu J_\mu^2-2C_\nu J_\nu-2C_\mu J_\mu=\kappa \quad , \quad (15)$$

where the K_i and C_i, $i=\nu,\mu$, in (15) are defined by

$$K_i=\frac{1}{\sigma_{Ti}^2}-\frac{1}{\sigma_{Ci}^2}, \quad C_i=\frac{\bar{J}_{Ti}}{\sigma_{Ti}^2}-\frac{\bar{J}_{Ci}}{\sigma_{Ci}^2}, \quad i=\nu,\mu$$

and

$$\kappa=\gamma-\left(\frac{\bar{J}_{T\nu}}{\sigma_{T\nu}^2}+\frac{\bar{J}_{C\nu}}{\sigma_{C\nu}^2}+\frac{\bar{J}_{T\mu}}{\sigma_{T\mu}^2}+\frac{\bar{J}_{C\mu}}{\sigma_{C\mu}^2}\right) .$$

The cross-section determined by (15) is then clearly an ellipse when

$$\sigma_{T\nu}>\sigma_{C\nu} \text{ and } \sigma_{T\mu}>\sigma_{C\mu} \quad ,$$

a parabola when either

$$\sigma_{T\nu} = \sigma_{C\nu} \text{ or } \sigma_{T\mu} = \sigma_{C\mu}$$

but not both, an hyperbola if either

$$\sigma_{T\nu} > \sigma_{C\nu} \text{ and } \sigma_{T\mu} < \sigma_{C\mu} \text{ or } \sigma_{T\nu} < \sigma_{C\mu} \text{ and } \sigma_{T\mu} > \sigma_{C\mu} ,$$

and a straight line if

$$\sigma_{T\nu} = \sigma_{C\nu} \text{ and } \sigma_{T\mu} = \sigma_{C\mu} .$$

A case of particular interest is that for which the covariance matrices $\underset{\sim}{M}_T$ and $\underset{\sim}{M}_C$ are identical. When this is true (10) becomes

$$-2\Delta\bar{\underset{\sim}{J}}^t \underset{\approx}{M}^{-1} \underset{\sim}{J} = \gamma + \bar{\underset{\sim}{J}}_C^t \underset{\approx}{M}^{-1} \bar{\underset{\sim}{J}}_C - \bar{\underset{\sim}{J}}_T^t \underset{\approx}{M}^{-1} \bar{\underset{\sim}{J}}_T , \qquad (16)$$

where

$$\Delta\bar{\underset{\sim}{J}} = \bar{\underset{\sim}{J}}_T - \bar{\underset{\sim}{J}}_C .$$

and $\underset{\approx}{M}$ is the common covariance matrix. Equation (16) has the form

$$\underset{\sim}{W}^t \underset{\sim}{J} = \sum_{n=1}^{N} W_n J_n = \text{constant}, \qquad (17)$$

where $\underset{\sim}{W}$ is the vector given by

$$\underset{\sim}{W} = \underset{\approx}{M}^{-1} \Delta \bar{\underset{\sim}{J}} . \qquad (18)$$

The constant on the right side of (17) is to be determined by the CFAR condition (11). This suggests that a change of variables such that one of the new variables J is given by

$$J = \sum_{n=1}^{N} W_n J_n \tag{19}$$

might be useful. Then the PFA given by (11) and the PTD given by (13) will depend only on the respective C and T marginal probability distributions for J.

According to the theorem, cited in Chapter II, concerning the effect of a linear transformation on a multivariate Gaussian probability distribution, the two probability distributions for J are univariate Gaussian with means $\bar{J}_i$, i = T, C, given by

$$\bar{J}_i = \underset{\sim}{W}^t \underset{\sim}{\bar{J}}_i = \sum_{n=1}^{N} W_n \bar{J}_{in}, \quad i = T, C \tag{20}$$

and a common variance σ^2 given by

$$\sigma^2 = \underset{\sim}{W}^t \underset{\approx}{M} \underset{\sim}{W} \quad . \tag{21}$$

The optimum target discrimination result for the univariate case derived in Appendix A now applies. That is, with ν defined by the CFAR condition

$$\frac{1}{\sqrt{2\pi}} \int_{\nu}^{\infty} e^{-\frac{x^2}{2}} dx = \phi \quad , \tag{22}$$

if $\Delta\bar{J} > 0$, then a target is declared to be present when

$$J > \bar{J}_C + \sigma\nu \quad ; \tag{23}$$

if $\Delta\bar{J} < 0$, then a target is declared to be present when

$$J < \bar{J}_C - \sigma\nu \quad . \tag{24}$$

B. LINEAR FILTERS

As observed in Section A, the optimum CFAR rule for target detection is non-linear, in fact quadratic, unless the covariance matrices $\underset{\approx}{M}_C$ and $\underset{\approx}{M}_T$ are identical. However, most detection algorithms for spatial, or for that matter temporal, channels are based on thresholding after the application of a linear filter.

The simplest example is, perhaps, the temporal filter that is sometimes referred to as Moving Target Identification (MTI), for which the basic idea is to detect a target's motion relative to what is presumed to be a stationary background. It is usually proposed for a staring system.

In this connection a single temporal channel is a frame that consists of the radiance distribution over the entire scene at a given instant of time. A sequence of such frames constitutes a set of temporal channels, just as an array of pixels constitutes a set of spatial channels.

The first-order MTI filtering process, the first difference, consists of subtracting one of two successive frames from the other. If a moving target is present but the background is fixed, this difference will be zero everywhere except at the two target positions, one in each frame.

The major problem encountered by MTI is the difficulty of maintaining registration for the background from one frame to

the next. Any motion of the sensor will cause an apparent movement in the background.

From one point of view this is a problem of correcting platform instability. However, for motion that is slowly varying or smooth (as distinguished, for example, from jitter) more complex temporal filtering may reduce or eliminate the error. Common filters for this purpose are second- or higher-order differences.

All such filters are linear and, in fact, are special cases of a sliding window weighted average, also known as a convolution. The general sliding window temporal filter is a linear transformation of the form

$$J'_\nu = \sum_{n=-\frac{(N-1)}{2}}^{\frac{(N-1)}{2}} W_n \, J_{n+\nu} \quad , \tag{25}$$

from a sequence of radiance values J_ν at an arbitrary pixel common to each frame to the sequence J'_ν. After the transformation is applied each term of the new sequence is usually thresholded and averaged, or averaged and thresholded, to form a simple spatial distribution which can then be processed further, as a spatial scene, to detect and locate possible targets.

Sliding window spatial filters that operate on a two dimensional array of pixels rather than a one-dimensional sequence of frames define analogous convolution transformations:

$$J'_\nu = \sum_n W_n \, J_{n+\nu} \quad . \tag{26}$$

In (26) the subscripts are two-component vectors in accordance with the notation introduced in Chapter II. The sum is taken over all pixels in an N by N array, for which N is an odd integer and the two-component vector ν locates the pixel at the center of the array.

In discussing either (25) or (26) it is convenient to set $\nu = 0$, which is equivalent to choosing a particular coordinate system for the discussion. A simple way to represent particular examples of (25) or (26), one that has become conventional, at least for the two-dimensional case, is to use a mask consisting of the weights W_n ordered as in the sequence or as in the array.

Examples for the one-dimensional (temporal) case are:

(1) the first difference mask

$$(0, -1, 1),$$

(2) the second difference mask

$$(1, -2, 1),$$

(3) the third difference mask

$$(0, -1, 3, -3, 1) \quad .$$

These masks apply to a sequence of discrete instants or time frames, which may be regarded as points along a temporal coordinate axis.

Examples for the two-dimensional (spatial) case are the so-called Laplacian filters:*

(1) the digital analogue of the Laplacian differential operator $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is represented by the mask

*Ref. 1, p. 482.

$$\begin{pmatrix} 0, & -1, & 0 \\ -1, & 4, & -1 \\ 0, & -1, & 0 \end{pmatrix},$$

(2) a rotationally symmetric modification by the mask

$$\begin{pmatrix} -1, & -1, & -1 \\ -1, & 8, & -1 \\ -1, & -1, & -1 \end{pmatrix},$$

(3) the digital analogue of the differential operator $\frac{\partial^4}{\partial x^2 \partial y^2}$ by the mask

$$\begin{pmatrix} 1, & -2, & 1 \\ -2, & 4, & -2 \\ 1, & -2, & 1 \end{pmatrix}.$$

These masks apply to a planar array of discrete pixels.

The Laplacian filters were designed to detect edges in a scene. The second filter, because it is rotationally symmetric, is actually a point detector and is therefore of particular interest for applications involving point targets.

Starting with the assumption that the spatial correlation function of the background distribution has the form (4) discussed in Chapter II, Ref. 10 derives the last filter as an approximation for the one that maximizes the signal-to-clutter ratio.

However, as observed in Chapter II, the correlation function (4) is that of an anisotropic background oriented fortuitously to conform with the track direction of the sensor as it scans the scene. If the same derivation were applied after assuming the isotropic correlation function (5) instead, a completely different type of filter, for which the continuous analogue would be a differentio-integral operator, would result.

Note that in each of the filter examples just presented the sum of the weights shown in the mask array is zero. Filters with this property are termed high pass because they completely eliminate a constant background distribution; i.e., they eliminate the DC component of the background distribution's spatial frequency expansion.

For $\nu = 0$, (25) and (26) both have the form

$$J' = \sum_n W_n J_n \quad , \tag{27}$$

except that the subscript n is a scalar in one case and a two-component vector in the other. If the mean vectors $\underset{\sim}{\bar{J}}_C$ and $\underset{\sim}{\bar{J}}_T$ and the covariance matrices $\underset{\approx}{M}_C$ and $\underset{\approx}{M}_T$ are associated with N-variate Gaussian probability distributions for the J_n in the case when targets are absent and in the case when a target is present, then according to the theorem of Chapter II the corresponding variables J' have univariate Gaussian probability distributions with means and variances given by

$$\bar{J}'_i = \sum_n W_n \bar{J}_{in}, \quad \sigma_i^2 = \underset{\sim}{W}^t \underset{\approx}{M}_i \underset{\sim}{W} = \sum_{n,m} W_n M_{inm} W_m, \quad i = C, T. \tag{28}$$

According to Appendix A and Section A of this chapter the optimum CFAR algorithm is non-linear except when the covariance matrices $\underset{\approx}{M}_T$ and $\underset{\approx}{M}_C$ are identical. Therefore, the use of a linear filter will be less than optimum unless this is, in fact, true.

When the two covariance matrices are identical a comparison of (20) and (21) with (28) shows that the weights corresponding to the linear filter that provides CFAR optimization will be the components of the vector $\underset{\sim}{W}$ given by (18). It is interesting to note that this filter is exactly the same as the one that Ref. 1 (pp. 560-561) shows will maximize the signal-

to-noise ratio if the signal power is identified with $(\bar{J}'_T - \bar{J}'_C)^2$ and the noise power with σ^2.*

For the statistical model proposed in Chapter II for spatial channels, when the target source exactly occupies a single pixel the covariance matrices $\underset{\sim}{M}_C$ and $\underset{\sim}{M}_T$ will never be equal, however. Since the procedure detailed in Appendix C for diagonalizing $\underset{\sim}{M}_C$ and $\underset{\sim}{M}_T$ simultaneously in this case is easily implemented, it may be simpler to use the true optimum CFAR detection algorithm, which is quadratic, than it would be to obtain what must necessarily be a sub-optimum linear filter.

Nevertheless, as observed in the discussion in Chapter II of the model applied to spectral channels, the two covariance matrices are at least proportional. If the proportionality constant is nearly equal to one, as is usually the case for spectral discrimination, and the magnitudes of the corresponding mean vectors are sufficiently different, the linear filter whose weights are given by (18) will provide near optimum CFAR discrimination. This follows from the fact that the quadratic term in (A-19) of Appendix A can then be neglected in comparison with the linear term.

C. TRACKING ALGORITHMS

For IR systems that detect point targets the false alarm rate is the specification that usually dominates the signal processing requirements. The desired rate may be as low as one per hour, implying false alarm probabilities as small as 10^{-10} whenever a target is declared.

Only an algorithm composed of a number of tests that are, in effect, guaranteed to be mutually independent has any hope of achieving so small a PFA. Such a guarantee may be possible for an algorithm based on temporal discrimination if the interval between successive time frames is sufficiently large. That is, the interval must be larger than any correlation time associated with spatial or spectral discriminants.

*Cf. also Ref. 10.

For a staring system, MTI differencing, as described in Section B, will tend to remove whatever correlated false alarms may result from preliminary target detection algorithms that are applied to the spatial or spectral channels. Scanning systems, on the other hand, generate false alarms that are spatially correlated when they are separated by less than a correlation distance associated with the background radiance distribution. One method of eliminating this kind of dependence has been to treat any group of detections that cluster so closely as a single detection located at the centroid of the group.

Most of the detections resulting from the CFAR algorithms will, of course, be false alarms. The final decision that a target is present will be referred to here as a target declaration to distinguish it from the CFAR detections established before this decision process is invoked.

Systems that are required to maintain very low false alarm rates usually rely upon tracking algorithms to provide target declarations. Those are algorithms that distinguish between target and clutter sources by means of the presumed trajectory characteristics of such sources when they are observed in motion over several time frames.

A tracking algorithm must deal with two types of trajectories: the non-accidental, which is due to the real motion of a source relative to the IR sensor, and the accidental, which is due to a random juxtaposition of clutter sources. Because the first type occurs in great variety, according to the scenario, the environment, and the system configuration, the effectiveness of an algorithm in dealing with it is difficult to evaluate except on a case-by-case basis.* However, it is

*Ref. 3 (pp. 310-330) describes a number of tracking algorithms that have been used for image processing. The list is far from exhaustive, however; recent IR system designs, for example, have introduced tracking algorithms based on trajectory characteristics that do not seem to have been exploited previously.

possible to estimate with some generality an algorithm's effectiveness in dealing with the second type of trajectory.

First of all, to the extent that the preliminary CFAR algorithms perform their assigned function, it may be assumed that every false alarm occurs with the same specified CFAR probability ϕ. Suppose that, in order to declare a target, the tracking algorithm requires the formation of some spatial pattern by a minimum of r detections, one from each of r different time frames. Suppose also that n is the total number of possible detection combinations that can form such a pattern. Then the probability that the algorithm will generate a false target declaration because of random false detections is given by

$$P = 1 - (1-\phi^r)^n \quad . \tag{29}$$

Suppose that the system's false alarm rate specification implies a probability of false target declaration no greater than P_o. Then from (29) it follows that r, which is the minimum number of detections required to establish a target track, will be determined by the inequality

$$r \geq \frac{\log \left[1- (1-P_o)^{1/n}\right]}{\log \phi} \quad . \tag{30}$$

Since the original reason for invoking the tracking algorithm was the premise

$$P_o << 1 \quad ,$$

(30) is essentially equivalent to

$$r \geq \frac{\log P_o - \log n}{\log \phi} \quad . \tag{31}$$

The right side of (31) approximates the right side of (30) with an error whose absolute value will certainly be less than 0.5; thus, the two inequalities will lead to the same bound when rounded off to the nearest integer.

The number n has a simple estimate which can be derived as follows. Suppose that each time frame contains a total of m pixels and that from one frame to the next each detection may be followed by a detection at any of k different pixels. If k is the same for each successive frame in the set that determines the admissible track, then

$$n = mk^{r-1} \quad . \tag{32}$$

If k varies it can be replaced by an average (geometric) value estimate, or, if the aim is to be conservative, by an upper bound.*

A substitution from (32) into (31) leads to

$$r \geq \frac{\log P_o - \log m - (r-1) \log k}{\log \phi} \quad ,$$

which, in turn, provides the result

$$r \geq \frac{\log P_o - \log m + \log k}{\log k + \log \phi} \quad , \tag{33}$$

provided that

$$k < \frac{1}{\phi} \quad . \tag{34}$$

Unless (34) is satisfied no positive value of r is possible. In that case the algorithm cannot meet the false alarm rate goal.

*It is certainly a tracker objective to make k a rapidly decreasing function of r.

The number k is a measure of the amount of branching permitted by the tracking algorithm, usually in order to allow for trajectory turns and an error tolerance. Therefore, the condition (34) implies that the complexity of the tracking algorithm will be limited by the CFAR specification that the preliminary detection algorithms are able to meet.

As an example, consider the case in which there are 10^7 pixels in the entire scene, the CFAR algorithms dispose of 99 percent of the background pixels, and it is required that the probability of a false target declaration be less than 10^{-10}. Then $m = 10^7$, $\phi = 0.01$, and $P_o = 10^{-10}$. If no branching is allowed, so that $k = 1$, according to (33) the number r of detections that must be considered in the target declaration algorithm's trajectory pattern before a target can be declared is greater than 8.5, i.e., 9 or more. If two branches are allowed, r must be 10 or more, if 3, 11 or more, and if 4, 12 or more.

Figure 7 contains a curve that depicts the lower bound on r as a function of k. As the figure indicates, k must be less than 100 because of the limitation imposed by (34).

Figure 7 also contains a second curve: for the case in which ϕ is equal to 0.001 (99.9% of the background pixels are eliminated by the preliminary detection algorithms) but the other parameters have the same values as in the first case. An inspection of this curve reveals that the increase in effectiveness of the detection algorithms permits a large increase in the number of allowed branches for a given number of detections in the track pattern. For example, for $\phi = 0.01$ no track with fewer than 9 pixels is satisfactory, and even if there are 9, only one branch is permitted; however, for $\phi = 0.001$ a track with as few as 6 pixels is adequate, and if there are 9 pixels as many as 17 branches are permitted.

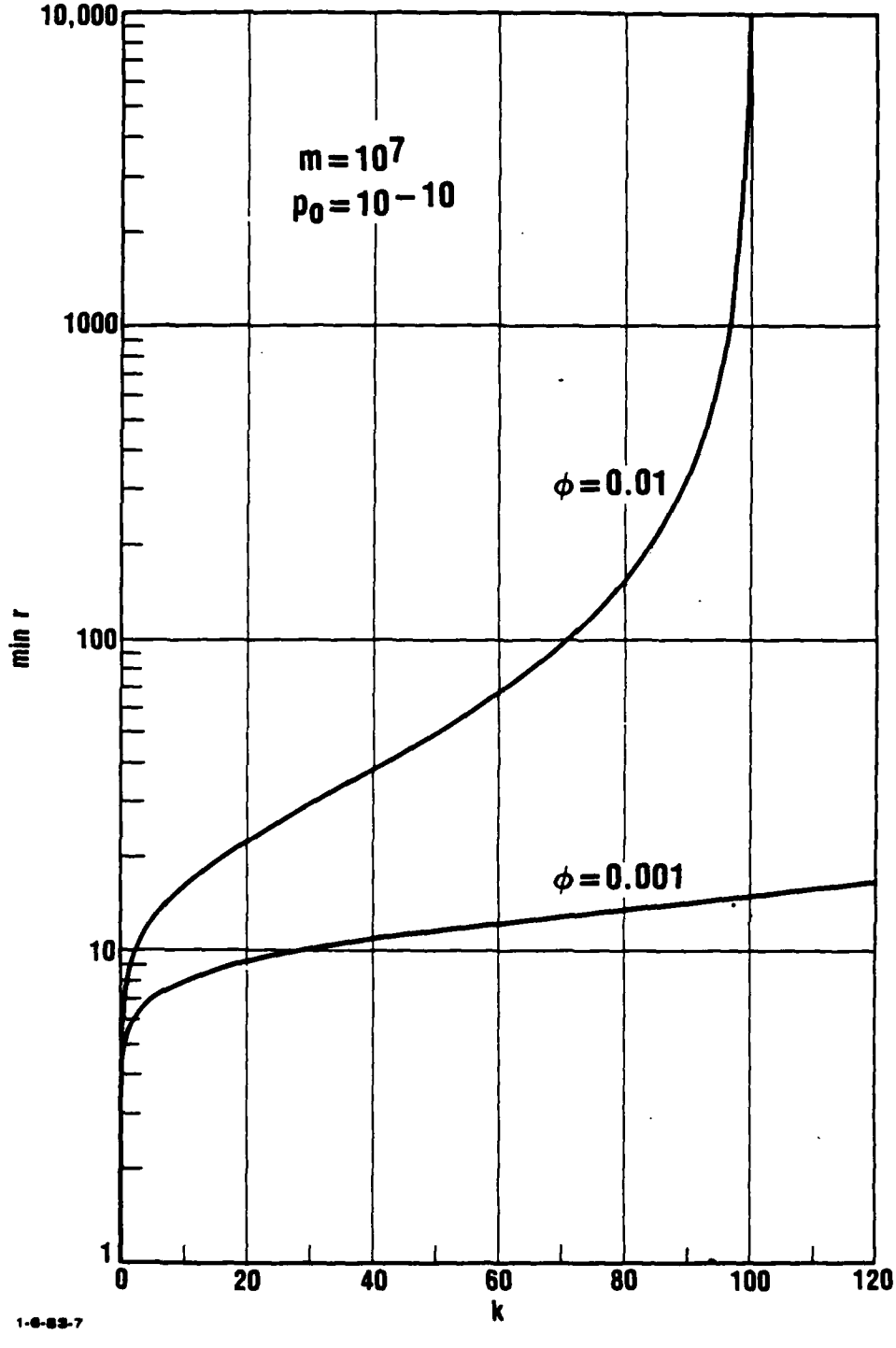


FIGURE 7. Minimum number of detections for a tracking algorithm versus number of branches

IV. AN EVALUATION OF TWO COMMON IR SIGNAL PROCESSING TECHNIQUES

A. INTRODUCTORY REMARKS

This chapter will discuss two unrelated techniques that are included in some IR signal processing approaches to target detection. The first technique, which is sometimes called background normalization (Ref. 12), is a method of setting a detection threshold that is adapted to the spatial variation of the background clutter. The second, which is incorporated in certain two- and three-color spectral discrimination algorithms, is a way of reducing the number of degrees of freedom in the data by using ratios of the spectral components rather than the components themselves.

The aim of the discussion will be to compare the effectiveness of the techniques with that of alternative approaches. The analysis that addresses this question here is actually an extension of the analysis in Appendices A and B of Ref. 4, which deal with the same topics in a more general way.

Appendix A of Ref. 4 characterizes background normalization in terms of an idealized version of the process. The present chapter will consider the specific process as it is ordinarily implemented.

Appendix B of Ref. 4 derives some general implications of the use of spectral component ratios in three-color systems. Here the concern will be with the probabilities of false alarm and target detection. For simplicity, the discussion will concentrate on two-color discrimination algorithms, although it is reasonable to suppose that the conclusions hold, at least qualitatively, for multi-spectral algorithms in general.

The treatment of both topics is self-contained in this paper. Nevertheless, there is not much overlap with the material in Ref. 4, which, therefore, might well furnish certain insights that the present discussion fails to provide.

B. BACKGROUND NORMALIZATION

Background normalization is a particular implementation of a general process called adaptive thresholding (cf. Ref. 13-18). The fundamental objective of signal processing, of course, is to set a detection threshold that is high enough to reject background clutter but low enough to pass a target signal. When the threshold selection varies with the local background distribution, i.e., is spatially adaptive, under CFAR conditions the target detection probability can be made larger than would be possible if the threshold were fixed for a whole scene.

Background normalization is essentially a method of estimating the clutter that would be observed at a given point P in the absence of a target. This estimate then provides a basis for setting a separate CFAR threshold for each point in the scene.

The prescribed estimate is just the average of the radiance, or of some function of the radiance (e.g., its square), measured at points surrounding P in a symmetrical pattern. For the scene as a whole the process amounts to a transformation of the background distribution by means of a sliding window average, which is a special case of the linear transformations discussed in Chapter III.

A simple example is the transformation defined by the mask

$$\begin{pmatrix} 1/16, & 1/16, & 1/16, & 1/16, & 1/16 \\ 1/16, & 0, & 0, & 0, & 1/16 \\ 1/16, & 0, & 0, & 0, & 1/16 \\ 1/16, & 0, & 0, & 0, & 1/16 \\ 1/16, & 1/16, & 1/16, & 1/16, & 1/16 \end{pmatrix} .$$

The point P corresponds to the central pixel in the window, and its eight neighbors are reserved as a guard against a possible overlapping signal from a target source that might have more spatial extent than was anticipated.

It is convenient in discussing the general case to introduce a Cartesian coordinate system chosen so that the point P is located at the origin; i.e., P will always have the coordinates (0,0). It will be assumed that the coordinates (x,y) of any other pixel in the window are integral multiples of a fixed quantity Δx in the horizontal direction and a fixed quantity Δy in the vertical direction.* Then, in an m by n window the pixels will be located at the points (x_ν, y_μ), for which

$$x_\nu = \nu\Delta x, \ \frac{1-n}{2} \leq \nu \leq \frac{n-1}{2} \ ,$$
$$y_\mu = \mu\Delta y, \ \frac{1-m}{2} \leq \mu \leq \frac{m-1}{2} \ . \qquad (35)$$

If the continuous background spatial distribution is given by a function S(x,y), the measured radiance (or a given function of the radiance) at each point (x_ν, y_μ) will be $S(x_\nu, y_\mu)$ in the absence of a target. Then background normalization consists of the assignment

$$\bar{S} = \frac{1}{M} \sum_{\nu,\mu} S(x_\nu, y_\mu) \ , \qquad (36)$$

where the sum is taken over a particular set of M points (x_ν, y_μ) out of the m n points in the window.

The set must satisfy just two conditions: (1) P is not a member; i.e., if (x_ν, y_μ) is in the set then either $x_\nu \neq 0$ or

*Usually Δx and Δy will be the same, but there is no particular advantage in assuming this restriction here.

$y_\mu \neq 0$; (2) the points that are members are located symmetrically with respect to P; i.e., if (x_ν, y_μ) is in the set then so is $(x_{-\nu}, y_{-\mu})$. It is evident from the second condition that

$$\bar{x} = \frac{1}{M} \sum_\nu x_\nu = 0 \text{ and } \bar{y} = \frac{1}{M} \sum_\mu y_\mu = 0 \quad , \tag{37}$$

where the barred quantities are averages of the indicated coordinates, calculated with respect to all points in the set.

The assignment (36) amounts to an interpolation of the background distribution to the point P from measured values observed at the M selected points (x_ν, y_μ). It was demonstrated in Appendix A of Ref. 4 that background normalization is consistent with a power series approximation that is valid, in general, up to the linear order. Therefore, it is natural to ask how it compares with an optimum linear interpolation from the given data.

An obvious choice for the comparison would be the estimate obtained from the linear function

$$\hat{S}(x,y) = ax + by + c \tag{38}$$

that fits the given data with the least square error. That is, the coefficients a, b, and c in (38) are to be determined from the condition that

$$\varepsilon = \sum_{\nu,\mu} [a\, x_\nu + b\, y_\mu + c - S(x_\nu, y_\mu)]^2 \tag{39}$$

be a minimum, where again the sum is taken over the M sample data points.

The standard method of calculating the coefficients, i.e., differentiating ε with respect to a, b, and c separately and

setting each derivative equal to zero, leads to the system of equations

$$\sum_{\nu,\mu} x_\nu \, [a \, x_\nu + b \, y_\mu + c - S(x_\nu, y_\mu)] = 0 \ ,$$

$$\sum_{\nu,\mu} y_\mu \, [a \, x_\nu + b \, y_\mu + c - S(x_\nu, y_\mu)] = 0 \ , \qquad (40)$$

$$\sum_{\nu,\mu} [a \, x_\nu + b \, y_\mu + c - S(x_\nu, y_\mu)] = 0 \ .$$

Because of (37) the last equation reduces to

$$c = \frac{1}{M} \sum_{\nu,\mu} S(x_\nu, y_\mu) \quad . \qquad (41)$$

But according to (38) the linear interpolation for the background at P is given by

$$\hat{S}(0, 0) = c \quad . \qquad (42)$$

A comparison of (36), (41), and (42) shows that the least-square-error linear interpolation for the radiance (or a given function of the radiance) at P is identical with the estimate given by background normalization.

Since there are only three parameters to be determined for the linear fit indicated by (38), it can be accomplished as long as there are more than three measured values of S(x,y). Generally, there will be more--e.g., eight in the case of a 3 x 3 element window, or at least sixteen in the case of a 5 x 5 element window.

This suggests the possibility of improving the background normalization technique by using an interpolation based on a square error fit of the data to a quadratic instead of a linear polynomial. That is, (38) would be replaced by

$$\hat{S}(x,y) = c + ax + b\,y + A_{11}\,x^2 + 2\,A_{12}\,xy + A_{22}\,y^2 \ , \quad (43)$$

and the coefficients c, a, b, A_{11}, A_{12}, A_{22} would be determined so as to minimize the square error

$$\varepsilon = \sum_{\nu}\sum_{\mu} [\hat{S}(x_\nu, y_\mu) - S(x_\nu, y_\mu)]^2 \ . \quad (44)$$

The resulting value of c, once again, would be the least square error estimate S(0, 0) of P.

The argument (based on the symmetrical distribution of data points about P) used to obtain (37) also implies that

$$\overline{x^3} = \overline{y^3} = 0 \ . \quad (45)$$

It will be found, as a result, that, on setting the derivatives of ε with respect to each of the six coefficients in (43) equal to zero, only three of the six least-square-error equations for the coefficients will contain c. Those equations are, in fact,

$$\begin{aligned} c + \overline{x^2}\,A_{11} + \overline{y^2}\,A_{22} &= S \ , \\ \overline{x^2}\,c + \overline{x^4}\,A_{11} + \overline{x^2y^2}\,A_{22} &= \overline{x^2S} \ , \\ \overline{y^2}\,c + \overline{x^2y^2}\,A_{11} + \overline{y^4}\,A_{22} &= \overline{y^2S} \ , \end{aligned} \quad (46)$$

where all barred quantities are averages over the M sample data points, e.g.,

$$\overline{x^2 S} = \frac{1}{M} \sum_{\nu,\mu} x_\nu^2 \, S(x_\nu, y_\mu) \quad .$$

If the coefficient determinant Δ of (46) is different from zero, Cramer's rule will provide an explicit solution for c, A_{11}, and A_{22}. However, only c is of interest here. It is given by

$$c = F \, \overline{S} - G \, \overline{x^2 S} - H \, \overline{y^2 S} \quad , \tag{47}$$

where

$$F = \frac{\overline{x^4}\,\overline{y^4} - \overline{x^2 y^2}^2}{\Delta} \quad ,$$

$$G = \frac{\overline{x^2}\,\overline{y^4} - \overline{y^2}\,\overline{x^2 y^2}}{\Delta} \quad , \tag{48}$$

$$H = \frac{\overline{y^2}\,\overline{x^4} - \overline{x^2}\,\overline{x^2 y^2}}{\Delta} \quad ,$$

and Δ is given by the determinant

$$\Delta = \begin{vmatrix} 1, & \overline{x^2}, & \overline{y^2} \\ \overline{x^2}, & \overline{x^4}, & \overline{x^2 y^2} \\ \overline{y^2}, & \overline{x^2 y^2}, & \overline{y^4} \end{vmatrix} \quad . \tag{49}$$

Since c is the least square error estimate of S(0, 0), it is evident from the form of (47) that the least-square-error estimate S_2 of quadratic order (replacing the linear-order estimate $S_1 = \overline{S}$) will be a weighted average; i.e.,

$$S_2 = c = \sum_{\nu,\mu} W_{\nu\mu} \, S(x_\nu, y_\mu) \quad . \tag{50}$$

The weights $W_{\nu\mu}$, obtained from an inspection of (47), are given by

$$W_{\nu\mu} = \frac{F\text{-}G\ x_\nu^2 - H\ y_\mu^2}{M} \tag{51}$$

$$= \frac{F\text{-}G\ \nu^2\ \Delta x^2 - H\ \mu^2\ \Delta y^2}{M} \quad .$$

As an example, consider the case, introduced earlier, of a 5 by 5 pixel window for which only the 16 border pixels are sample data points. For this case the ordinary background normalization, or linear, estimate consists of the average

$$S_1 = \sum_{\nu,\mu} W_{\nu\mu}\ S(x_\nu, y_\mu) \quad , \quad \nu = \pm 2 \text{ or } \mu = \pm 2 \quad ,$$

with equal weights,

$$W_{\nu\mu} = \frac{1}{16} \quad .$$

For the quadratic interpolation estimate it is necessary, first, to calculate the averages $\overline{x^2}$, $\overline{y^2}$, $\overline{x^4}$, $\overline{y^4}$, $\overline{x^2y^2}$, which can be done without much difficulty by using the mask introduced earlier as a guide. The results are

$$\overline{x^2} = \frac{5(x_2^2 + x_{-2}^2) + 2\ (x_1^2 + x_{-1}^2)}{16}$$

$$= \frac{5(4+4) + 2\ (1+1)}{16}\ \Delta x^2 = 2.75\ \Delta x^2 \quad ,$$

$$\overline{y^2} = 2.75\ \Delta y^2 \ ,$$

$$\overline{x^4} = 5(x_2^4+x_{-2}^4) + 2\ (x_1^4+x_{-1}^4) = 10.25\ \Delta x^4 \ ,$$

$$\overline{y^4} = 10.25\ \Delta y^4 \ ,$$

$$\overline{x^2y^2} = \frac{(x_1^2+x_{-1}^2)\ (y_2^2+y_{-2}^2) + (x_2^2+x_{-2}^2)\ (y_1^2+y_{-1}^2) + (x_2^2+x_{-2}^2)\ (y_2^2+y_{-2}^2)}{16}$$

$$= 6\ \Delta x^2\ \Delta y^2 \ .$$

The determinant Δ, defined by (49), is therefore given by

$$\Delta = 4.78125\ \Delta x^4\ \Delta y^4 \ .$$

Then the calculations indicated by (48) provide the results

$$\begin{aligned} F &= 14.44444 \ , \\ G &= \frac{2.44444}{\Delta x^2} \ , \\ H &= \frac{2.44444}{\Delta y^2} \ . \end{aligned} \tag{52}$$

Finally, the weights can be obtained by substituting from (52) into (51). The results are

$$\begin{aligned} W_{\pm 20} = W_{0\pm 2} &= \frac{F-4G\Delta x^2}{16} = 0.29167 \ , \\ W_{\pm 2\pm 1} = W_{\pm 1\pm 2} &= \frac{F-G\Delta x^2 - 4H\Delta y^2}{16} = 0.13889 \ , \\ W_{\pm 2\pm 2} &= \frac{F-4G\Delta x^2 - 4H\Delta y^2}{16} = -.31944 \ . \end{aligned} \tag{53}$$

The corresponding mask will be

$$\begin{pmatrix} -.31944, & 0.13889, & 0.29167, & 0.13889, & -.31944 \\ 0.13889, & 0, & 0, & 0, & 0.13889 \\ 0.29167, & 0, & 0, & 0, & 0.29167 \\ 0.13889, & 0, & 0, & 0, & 0.13889 \\ -.31944, & 0.13889, & 0.29167, & 0.13889, & -.31944 \end{pmatrix} .$$

C. MULTI-COLOR ALGORITHMS BASED ON SPECTRAL COMPONENT RATIOS

An N-color IR system collects data in N spectral channels defined by N distinct, non-overlapping wavelength bands which are presumably chosen because the spectral signatures that they provide for targets differ as much as possible from those that they provide for backgrounds. As is customary in this paper, N-component vectors $\underset{\sim}{J}_T$ and $\underset{\sim}{J}_C$ will represent the radiance distribution over the channels, the first for the case in which a target is present and the second for the case in which targets are absent.

Some two- and three-color target detection algorithms that have been proposed do not operate directly on the components J_i, $i=1,\ldots,N$ of the vectors $\underset{\sim}{J}_T$ and $\underset{\sim}{J}_C$, but rather on the ratios of N-1 of the J_i, $i=1,\ldots,N-1$, to a single component J_N. That is, the ratio variables

$$X_i = \frac{J_i}{J_N}, \quad i=1,\ldots,N-1 , \tag{54}$$

replace the variables J_i, $i=1,\ldots,N$, and it is the X_i that enter into the target detection algorithms. As a result, there must be a certain loss of information since the number of free variables will then be reduced by one. The question is: how does this use of ratios in a target detection algorithm affect the false alarm probability?

For simplicity the discussion will be confined to two-color systems, although similar conclusions may be expected in the case of systems that employ three or more colors. For two colors it is possible to construct a simple graphical representation of the pair of measured spectral components J_1 and J_2.

This is illustrated by Fig. 8 which depicts a planar coordinate system for points that are defined when J_1 and J_2 are regarded as cartesian coordinates. The figure represents a data plane in which every point corresponds to a pair of measurements in the two spectral bands of interest, and every pair of such measurements corresponds to a point in the plane.

Assume that there is a distinct bivariate joint probability distribution for (J_1, J_2) corresponding to the target source and another such distribution corresponding to the clutter. In terms of its probability distribution a mean point $(\bar{J}_1, \bar{J}_2)$ will be defined for the target and another will be defined for clutter. These are indicated by the labels "target" and "clutter" in the figure.

To each point in the data plane there is an associated line through the point and the origin of the coordinate system. The ratio of the corresponding spectral components will be equal to the slope of the line or the reciprocal of the slope, depending upon how the ratio is defined. Figure 8 shows the lines (solid) associated in this way with the target and clutter means.

A straightforward discrimination criterion is provided by the following rule.* If for a pair of measurements J_1 and J_2 the value of the probability density function (PDF) corresponding to the target is greater than the value of the PDF corresponding to clutter, the source is presumed to be the target. Otherwise, the source is presumed to be clutter.

*This rule is introduced here instead of one based on a CFAR requirement to simplify the calculations that illustrate the major points of interest. It also makes it possible to normalize the evaluation to one for which the false alarm probability is the figure of merit.

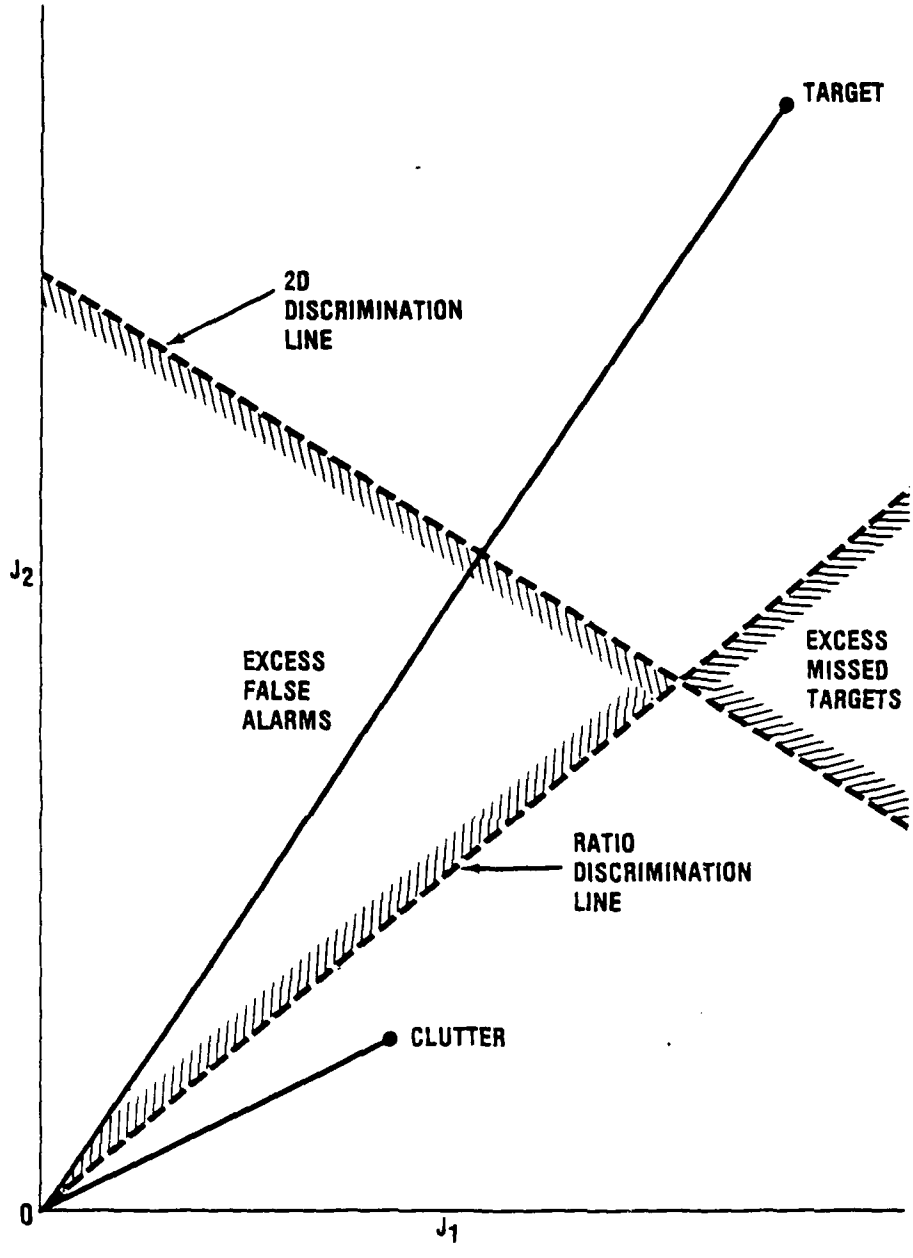


FIGURE 8. 2D versus ratio discrimination for 2-color systems

Then the curve defined by the equation that is formed when the target PDF is set equal to the clutter PDF divides the data plane into two regions: one consisting of points regarded as due to the target and the other consisting of points regarded as due to clutter. This boundary is indicated in Fig. 8 by the line labeled "2D discrimination line".

Although the boundary is shown in the figure as straight, in general it will be a curve or, in fact, it may even consist of two distinct branches of a curve. If the PDFs are both bivariate gaussian the boundary will be a conic section (Chapter III, Section A), i.e., a parabola, an ellipse, or a hyperbola. If the covariance matrix for the target PDF and that for the clutter PDF are identical, in the case of gaussian distributions the boundary will be a straight line. Appendix A contains a detailed discussion of these and related matters.

The target and clutter probability distributions will each induce a corresponding univariate distribution for the ratio of spectral components.* A discrimination criterion similar to that based on the bivariate PDFs can be formulated in terms of the ratio PDFs.

When the target and clutter ratio PDFs are set equal the solution of that equation provides a boundary between the region consisting of points designated as target and the region consisting of points designated as clutter by the ratio discrimination criterion. This boundary is a line or lines, with slopes given by the solution of the equation, passing through the origin of the coordinate system, and it defines regions that are angular sectors. This is illustrated in Fig. 8 by a line labeled "ratio discrimination line".

It is evident that the 2D discrimination rule and the ratio rule do not always agree. The region labeled "excess false alarms" in the figure consists of points that are false

*See Appendix D for a derivation of the PDF.

alarms from the point of view of the 2D rule, and the region labeled "excess missed targets" consists of points that are missed target detections from the same point of view. From the ratio rule point of view the same regions would be applicable with the labels reversed.

Figure 9 illustrates a similar data plane configuration, except that in this case the target and clutter means have the same ratio, although the mean points are still separated by a considerable margin. Note that the ratio boundary between designated target and clutter points consists of two lines in this example.

The triangular region labeled "ratio false alarms" consists of points that are false alarms from the point of view of the 2D rule. The two angular sector regions labeled "2D excess false alarms" consist of points that are false alarms from the point of view of the ratio rule. The actual false alarm probabilities are determined not by the areas of these regions but by the result of integrating the bivariate clutter PDF over the regions.

Figure 10 illustrates graphically several cases of a bivariate Gaussian distribution. The solid-line ellipse represents a curve of constant probability for the case of uncorrelated spectral components with the standard deviation of one component equal to ten times that of the other. Also, the ratio corresponding to the mean point is defined to be $\bar{J}_1/\bar{J}_2$ and is equal to 2/3.

The dashed-line ellipses in Fig. 10 are lines of constant probability for distributions that are obtained from the original distribution by rotating the coordinate system about the mean point through various angles as indicated. This provides cases in which the covariance matrix is not diagonal, i.e., in which the spectral components are correlated.

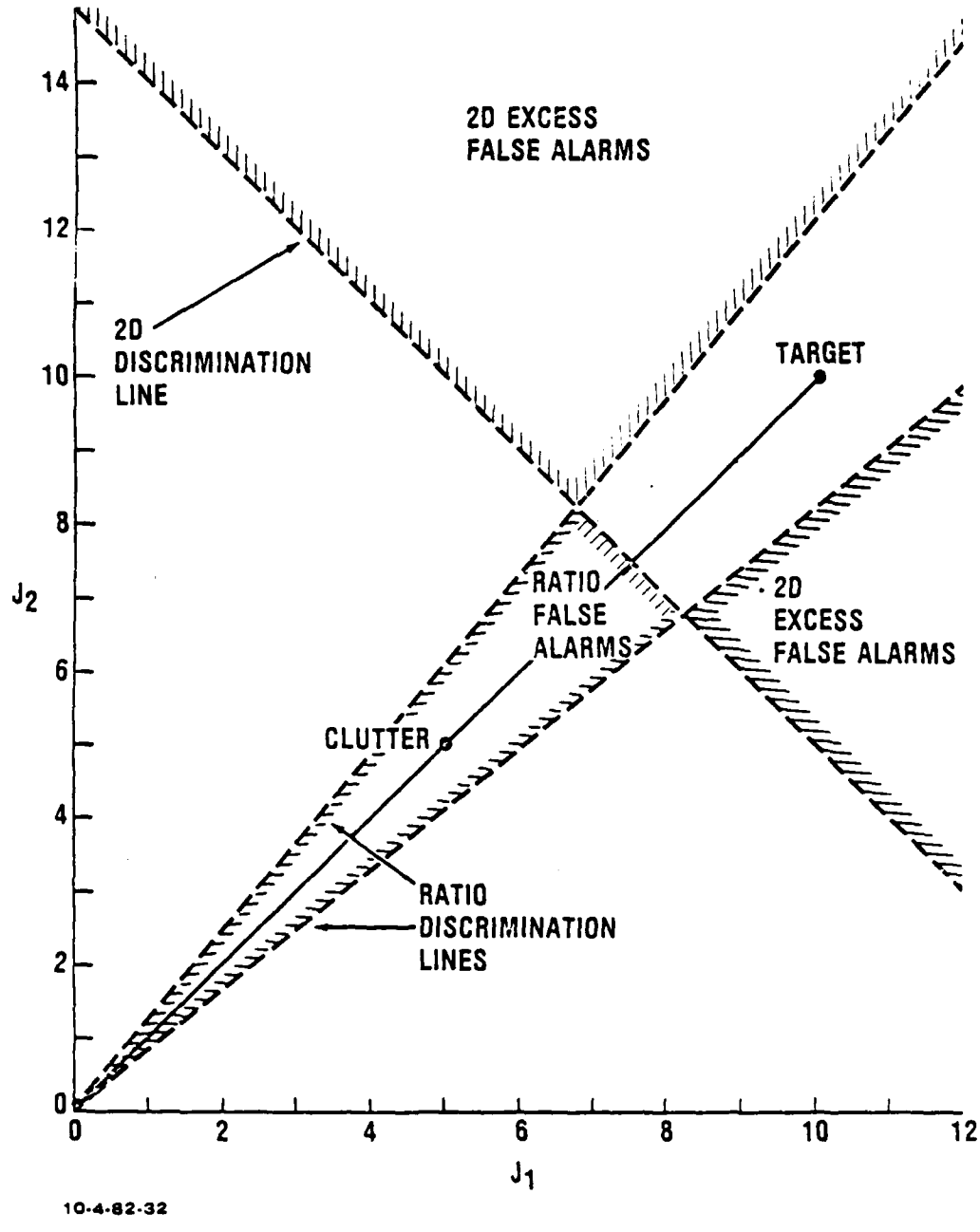


FIGURE 9. 2D versus ratio discrimination for 2-color systems

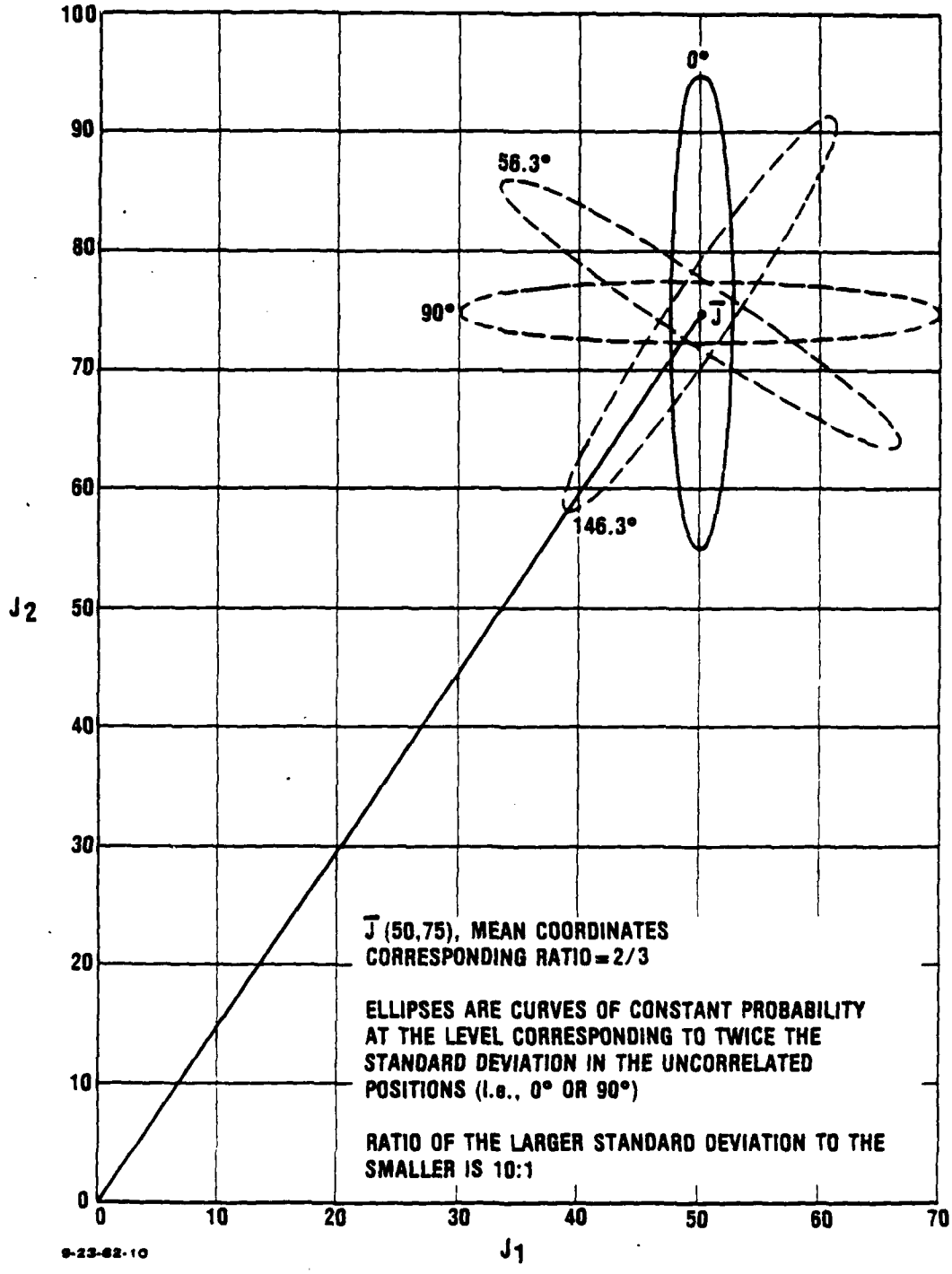


FIGURE 10. Bivariate probability densities with various amounts of correlation induced by rotation about the mean

Figure 11 shows curves that represent the ratio PDF* corresponding to each bivariate PDF illustrated in Fig. 10. Note that each curve has a single mode which occurs near, but not exactly at, the ratio of the mean components, i.e., 2/3. Also note, by comparison with Fig. 10, that the largest mode occurs for the case in which the major axis of the corresponding bivariate constant probability ellipse is colinear with the line joining the mean point and the origin of the coordinate system. Further, the smallest mode occurs for the case in which it is the minor axis that is colinear with that line.

To calculate the probability of false alarm (PFA) for either the ratio or the 2D rule, as observed in Chapter III, it is only necessary to integrate the clutter PDF over the appropriate region for a bivariate Gaussian distribution. The region will always be bounded by straight lines whenever the target and background covariance matrices are the same. According to the mathematical model proposed in Chapter II, this will generally be the case for spectral discrimination of point targets. Appendix D shows in detail how such integrals can be evaluated efficiently.

To make the false alarm probability calculation particularly easy, consider the simplest possible case, in which the target and clutter probability distributions are both bivariate Gaussian with uncorrelated spectral components having identical standard deviations. In accordance with the mathematical model of Chapter II the standard deviation will be the same for the target and clutter distributions, as well. For this case Table 1 provides false alarm probabilities due to the ratio rule and to the 2D (bivariate) rule for three different sets of means given in units of the common standard deviation.

*Equation D-14 of Appendix D was used to plot these curves.

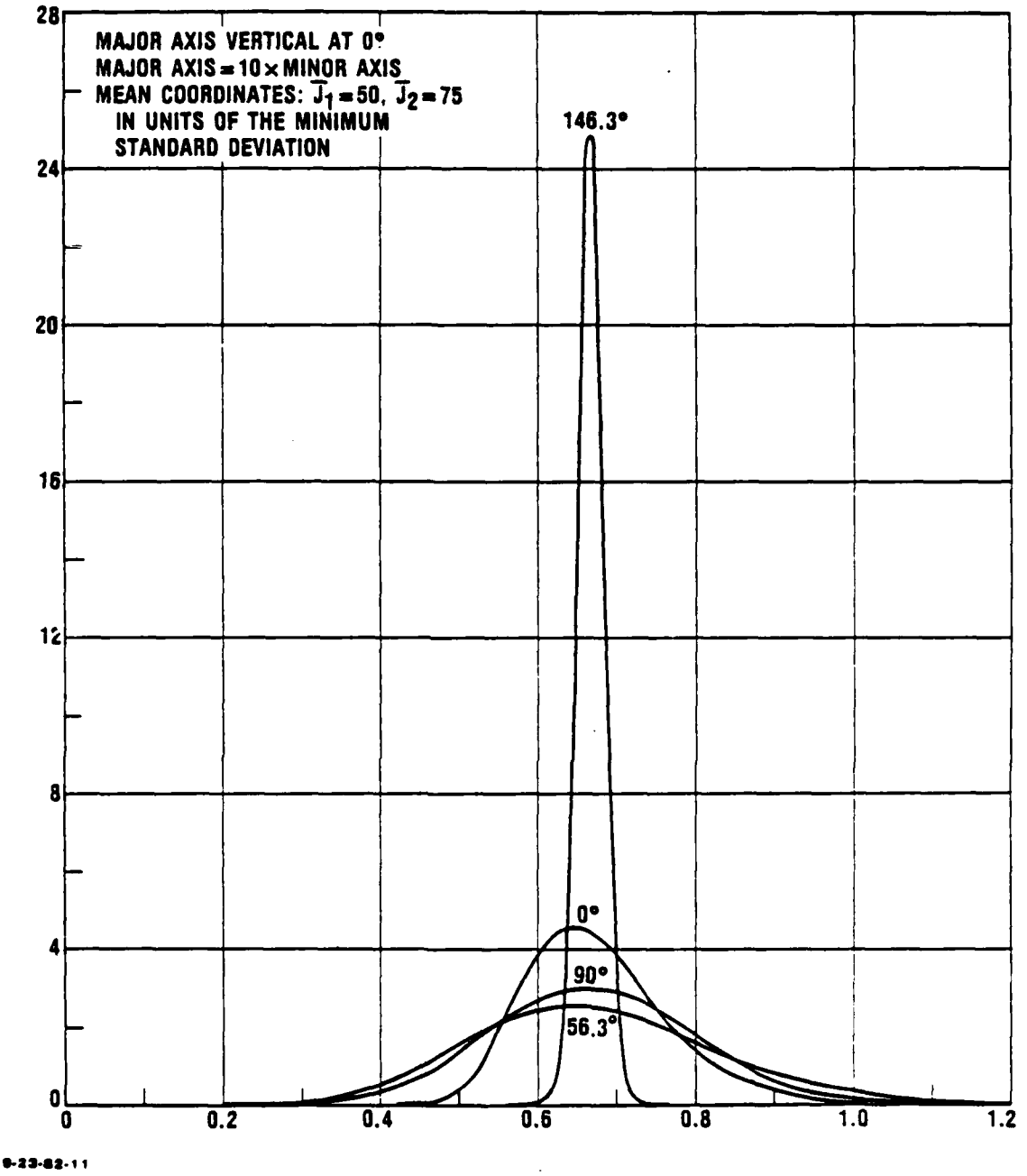


FIGURE 11. Probability densities for the ratio $\frac{J_1}{J_2}$ when J_1 and J_2 are uncorrelated at 0^0 and correlated by rotating distribution counterclockwise about the mean point $(\bar{J}_1, \bar{J}_2)$

TABLE 1. FALSE ALARM PROBABILITIES FOR RATIO AND 2D DISCRIMINATION RULES

ARTIFICIAL DATA

SPECTRAL BAND	CLUTTER MEAN	TARGET MEAN	STANDARD DEVIATION	CORRELATION COEFFICIENT	PFA	RULE
					0.5034	Ratio
$\bar{J}_1$	5	10	1	0	2×10^{-4}	2D
$\bar{J}_2$	5	10	1	0	0.5032	Ratio (excess)
					$< 10^{-4}$	2D (excess)
					0.4948	Ratio
$\bar{J}_1$	5	10	1	0	2×10^{-4}	2D
$\bar{J}_2$	6	11	1	0	0.4946	Ratio (excess)
					$< 10^{-4}$	2D (excess)
					0.0339	Ratio
$\bar{J}_1$	5	5	1	0	$< 10^{-4}$	2D
$\bar{J}_2$	5	13	1	0	0.0338	Ratio (excess)
					$< 10^{-4}$	2D (excess)

It is seen in the table that for the case in which the target and clutter means have the same ratio of spectral components, as illustrated in Fig. 9, the ratio rule produces a false alarm probability that is more than 50 percent, while the 2D rule's false alarm probability is about 0.02 percent. When the means are shifted slightly so that the target and clutter means are no longer associated with identical component ratios the ratio rule false alarm probability improves slightly to a little less than 50 percent while the 2D rule false alarm probability remains essentially the same. When the means are shifted by a greater amount so that the ratio associated with the target mean is somewhat greater than 2-1/2 times the ratio associated with the clutter mean the false alarm probability due to the ratio test improves considerably. However, it is still more than 3 percent, while the false alarm probability due to the 2D rule is less than 0.01 percent.

Tables 2, 3 and 4 contain the results of similar calculations based on data taken from Ref. 9 for natural terrain backgrounds. Data for the targets were made up by using equivalent temperature means that are 3σ or 5σ above the corresponding background means for one or both of the wavelength bands. One scene is a conifer forest in Michigan and the other is a mountainous area in Nevada.

An examination of the tables indicates that the ratio rule produces consistently higher false alarm probabilities than does the 2D rule. In many of the examples the PFAs differ by an order of magnitude or more.

TABLE 2. FALSE ALARM PROBABILITIES FOR RATIO AND 2D DISCRIMINATION RULES

Background Type: Mountainous Terrain (Nellis AF Base, Nevada)

Conditions: AM (1100, 2-26-78), high overcast, light haze, visibility 15 mi

Aircraft: Altitude 1000 ft, ground speed 200 ft/sec, flight direction East

Area Covered: 1750 ft x 6750 ft; Depression Angle: 35 deg; IFOV: 2.5 mrad

Radiance Units: deg k

SPECTRAL BAND	CLUTTER MEAN	TARGET MEAN	STANDARD DEVIATION	CORRELATION COEFFICIENT	PFA	RULE
					0.0453	Ratio
3.0-4.2 μ	285.68	303.93	3.65	0.539	2.2×10^{-3}	2D
4.5-5.5 μ	283.12	290.62	1.5		0.0440	Ratio (excess)
					9×10^{-4}	2D (excess)
					1.6×10^{-3}	Ratio
3.0-4.2 μ	285.68	303.93	3.65	0.539	1.5×10^{-3}	2D
4.5-5.5 μ	283.12	283.12	1.5		3×10^{-4}	Ratio (excess)
					2×10^{-4}	2D (excess)
					0.1174	Ratio
3.0-4.2 μ	285.68	285.68	3.65	0.539	1.5×10^{-3}	2D
4.5-5.5 μ	283.12	290.62	1.5		0.1165	Ratio (excess)
					7×10^{-4}	2D (excess)
					0.0388	Ratio
3.0-4.2 μ	285.68	296.63	3.65	0.539	0.0375	2D
4.5-5.5 μ	283.12	283.12	1.5		5.2×10^{-3}	Ratio (excess)
					3.9×10^{-3}	2D (excess)
					0.1545	Ratio
3.0-4.2 μ	285.68	296.63	3.65	0.539	0.0436	2D
4.5-5.5 μ	283.12	287.62	1.5		0.1276	Ratio (excess)
					0.0167	2D (excess)
					0.2381	Ratio
3.0-4.2 μ	285.68	285.68	3.65	0.539	0.0375	2D
4.5-5.5 μ	283.12	287.62	1.5		0.2166	Ratio (excess)
					0.0159	2D (excess)

TABLE 3. FALSE ALARM PROBABILITIES FOR RATIO AND 2D DISCRIMINATION RULES

Background Type: Mountainous Terrain (Nellis AF Base, Nevada)
Conditions: AM (0930, 2-25-78), high thin scattered clouds, visibility 15 mi
Aircraft: Altitude 1750 ft, ground speed 200 ft/sec, flight direction West
Area Covered: 1750 ft x 6750 ft; Depression Angle: 90 deg; IFOV: 2.5 mrad
Radiance Units: deg k

SPECTRAL BAND	CLUTTER MEAN	TARGET MEAN	STANDARD DEVIATION	CORRELATION COEFFICIENT	PFA	RULE
					$< 10^{-4}$	Ratio
3.0-4.2 μ	291.12	355.57	8.89	0.894	$< 10^{-4}$	2D
4.5-5.5 μ	283.73	307.08	4.67		$< 10^{-4}$	Ratio (excess)
					$< 10^{-4}$	2D (excess)
					$< 10^{-4}$	Ratio
3.0-4.2 μ	291.12	355.57	8.89	0.894	$< 10^{-4}$	2D
4.5-5.5 μ	283.73	283.73	4.67		$< 10^{-4}$	Ratio (excess)
					$< 10^{-4}$	2D (excess)
					0.0144	Ratio
3.0-4.2 μ	291.12	291.12	8.89	0.894	$< 10^{-4}$	2D
4.5-5.5 μ	283.73	307.08	4.67		0.0144	Ratio (excess)
					$< 10^{-4}$	2D (excess)
					0.1180	Ratio
3.0-4.2 μ	291.12	317.79	8.89	0.894	0.0616	2D
4.5-5.5 μ	283.73	297.74	4.67		0.0758	Ratio (excess)
					0.0194	2D (excess)
					0.0911	Ratio
3.0-4.2 μ	291.12	291.12	8.89	0.894	4×10^{-4}	2D
4.5-5.5 μ	283.73	297.74	4.67		0.0909	Ratio (excess)
					2×10^{-4}	2D (excess)
					3.4×10^{-3}	Ratio
3.0-4.2 μ	291.12	317.79	8.89	0.894	4×10^{-4}	2D
4.5-5.5 μ	283.73	283.73	4.67		3.2×10^{-3}	Ratio (excess)
					2×10^{-4}	2D (excess)

TABLE 4. FALSE ALARM PROBABILITIES FOR RATIO AND 2D DISCRIMINATION RULES

Background Type: Conifer Forest (Michigan)

Conditions: 1230 (Winter, 4-3-79, 4-4-79), *no clouds, snow-covered ground, air* temperature - 2 deg C

Aircraft: Altitude 1750 ft, ground speed 202 ft/sec, flight direction NNW

Area Covered: 1650 ft x 1750 ft; Depression Angle: 90 deg; IFOV: 2.5 mrad

Radiance Units: deg k

SPECTRAL BAND	CLUTTER MEAN	TARGET MEAN	STANDARD DEVIATION	CORRELATION COEFFICIENT	PFA	RULE
					0.0190	Ratio
3.5-3.9 μ	281.77	300.115	3.6689	0.169	5×10^{-4}	2D
4.5-5.5 μ	277.58	280.751	0.6341		0.0187	Ratio (excess)
					2×10^{-4}	2D (excess)
					5.6×10^{-3}	Ratio
3.5-3.9 μ	281.77	300.115	3.6689	0.169	5.6×10^{-3}	2D
4.5-5.5 μ	277.58	277.58	0.6341		1×10^{-4}	Ratio (excess)
					1×10^{-4}	2D (excess)
					0.3322	Ratio
3.5-3.9 μ	281.77	281.77	3.6689	0.169	5.6×10^{-3}	2D
4.5-5.5 μ	277.58	280.751	0.6341		0.3293	Ratio (excess)
					0.0027	2D (excess)
					0.1064	Ratio
3.5-3.9 μ	281.77	292.777	3.6689	0.169	0.0250	2D
4.5-5.5 μ	277.58	279.482	0.6341		0.0910	Ratio (excess)
					0.0095	2D (excess)
					0.0643	Ratio
3.5-3.9 μ	281.77	292.777	3.6689	0.169	0.0641	2D
4.5-5.5 μ	277.58	277.58	0.6341		6×10^{-4}	Ratio (excess)
					4×10^{-4}	2D (excess)
					0.4058	Ratio
3.5-3.9 μ	281.77	281.77	3.6689	0.169	0.0643	2D
4.5-5.5 μ	277.58	279.482	0.6341		0.3710	Ratio (excess)
					0.0295	2D (excess)

V. SUMMARY AND CONCLUSIONS

A. SUMMARY OF TOPICS COVERED

Based on the assumption that IR measurement data separated into N distinct channels have an N-variate Gaussian probability distribution, this paper formulates a mathematical model for the background radiance in the presence and in the absence of targets. The formulation includes both spectral and spatial discriminants for the case of point targets.

According to the model as proposed, if the N data channels are defined as spectral bands it is usually the case that the probability distributions associated with the presence of a target and with the absence of any target differ significantly only in their N dimensional mean vectors. That is, their N by N covariance matrices are assumed to be nearly identical. This will be true as long as the target occults only a small fraction of the sensor's footprint.

On the other hand, if the N data channels are defined in terms of the spatial discriminant, i.e., so that each channel represents the radiance level at a single pixel in an N pixel window, the covariance matrices associated with the presence or absence of a target will differ unless the target occupies just a small fraction of a pixel. In fact, the model assumes an explicit form (Chapter II, Section E) for the difference of the two matrices when the target exactly fills a single pixel.

For certain calculations it is convenient to change coordinates by means of a principal axis transformation relative to the covariance matrix associated with an N-variate Gaussian probability distribution. Appendix C shows in detail how this

can be done simultaneously for the two covariance matrices associated with spatial data channels in the presence and in the absence of a target.

The purpose of the model is to provide a means for obtaining rough evaluations of proposed target discrimination schemes on the basis of what may be regarded as a minimal acceptance standard. The analysis (Chapter III) covers optimum CFAR discrimination and also includes a consideration of the effectiveness of tracking algorithms (Chapter III, Section C) after CFAR discrimination algorithms have been applied.

In addition to these topics and some related detail on linear filtering (Chapter III, Section B) and how to calculate various quantities of interest, this paper also deals (Chapter IV) with two special subjects. One is a method of adaptive thresholding known as background normalization. The other is the question of whether it is useful or harmful for multi-color systems to use ratios of spectral components, rather than the components themselves, in target discrimination processing.

The ordinary background normalization process amounts to a linear least-square-error interpolation of local measurement data to predict the value of the background radiance, or some function of the radiance, in a given direction in the absence of a target. This paper shows how to extend the interpolation to include terms of quadratic order by means of a special linear filter. As an example, the weights that define the filter mask for the case of a 5 by 5 pixel window are calculated.

The analysis required for the multi-color question involves a calculation of the probability distribution for spectral component ratios. Properties of the corresponding probability density for the two-color case are discussed in detail.

Chapter IV contains tables of false alarm probability calculations for the two-color case for two discrimination algorithms, one based on the one-dimensional distribution for the ratio of

the two spectral components and the other on the two-dimensional bivariate Gaussian distribution for the components themselves. The tables answer the question concerning the relative merit of the two approaches. Most of the data used for the calculations are taken from Ref. 9, which provides results of radiance measurements over several wavelength bands for a variety of terrain backgrounds.

B. CONCLUSIONS

(1) Experimental data (Ref. 9) for a variety of terrain backgrounds, especially those unaffected by human intervention, exhibit radiance distributions that are well approximated (out to 2σ, 3σ or more) by Gaussian probability density functions. This may be adequate for realistically estimating the effect of preliminary detection algorithms for which the false alarm rate requirements are relatively modest. However, for some scenes that have been affected by human intervention, notably A.P. Hill, Virginia and Baltimore, Maryland, the approximation is poor. At any rate, the assumption of a Gaussian distributed background provides a minimal standard against which to measure an algorithm's clutter rejection performance.

(2) The data in Ref. 9, provided by the Environmental Research Institute of Michigan (ERIM), is presented in a form that is well-suited to mathematical modeling of the spectral distribution of terrain background radiance. It is also useful for testing hypotheses concerning the spatial distribution of the radiance. Data for other types of background, e.g., clouds, would be similarly useful if gathered and presented in the same form.

(3) The ERIM data supports the assumption that a natural (e.g., a conifer forest) scene is spatially isotropic but not the assumption that the cross-correlation function is exponential. In fact, the Wiener spectra given in Ref. 9 for a conifer forest are not consistent with any simple power law generalization

of the spectrum associated with an exponential cross-correlation function. The isotropic character of the spatial distribution obtained for a conifer forest in the ERIM data contradicts a model that is sometimes assumed (cf. Ref. 10) for the cross-correlation.

(4) If the distribution of IR background radiance over N channels (spectral, spatial or temporal) is N-variate Gaussian when targets are present and when they are not, the optimum CFAR target discrimination criterion is an inequality involving a quadratic function of the measured data unless the covariance matrix for the background in the absence of any target is identical with that for the background when a target is present.

(5) When the two covariance matrices are identical the optimum discrimination algorithm is equivalent to applying a linear digital filter and then thresholding. This optimum linear filter is the same as the well-known filter that maximizes signal-to-noise if the signal power is identified with the square of the difference between the mean target and mean background signals and the noise power is identified with the variance of the background distribution.

(6) For spectral channels the two covariance matrices approach equality when the target occults a small fraction of the sensor's footprint, as is usually the case. For spatial channels the two covariance matrices approach equality when the target occupies just a small fraction of a pixel. The optimum filter will be approximately linear if this happens, particularly if the mean target signal differs from the mean background signal by several standard deviations. When the target size is of the order of a pixel the spatial covariance matrices will differ, and the optimum spatial filter will not be linear.

(7) If a tracking algorithm is used for the final decision whether a target is or is not present after the application of one or more preliminary CFAR detection algorithms has eliminated

most of the candidate detections, the effectiveness of the tracking algorithm will depend upon the effectiveness of the preliminary algorithms. In fact, unless the false alarm probability after the preliminary detection phase is below a certain critical value, no tracking algorithm can satisfy a given false alarm rate requirement. Moreover, the sensitivity of a tracking algorithm to error or to unpredicted target accelerations will increase rapidly with an increase in the false alarm probability for the preliminary detection phase.

(8) When ratios of spectral components are used by multicolor systems to discriminate between targets and background rather than the components, themselves, the discrimination algorithm will be less effective. In particular, for the two-color case applied to typical natural background data obtained by ERIM (Ref. 9), when an algorithm based on the ratio of the two spectral components is used instead of one based on the two-dimensional distribution of the components the calculated false alarm probability is consistently larger, often by an order of magnitude or more.

REFERENCES

1. W.K. Pratt," Digital Image Processing," Wiley, New York, 1978.

2. M.R. Wohlers, et al., "Narrow Band Infrared Detection," Aerodyne Research, Inc. (Center for Computer Applications, Bedford, Mass.) Final Report # AFWAL-TR-81-1087 for the Avionics Laboratory, AFWAL, AFSC, Wright-Patterson AFB, Ohio, July 1981 (SECRET).

3. A. Rosenfeld and A.C. Kak, "Digital Picture Processing," Academic Press, New York, 1976.

4. M. Burns and I. Kay, "Aspects of Signal Processing in Airborne Infrared Search and Track Systems," IDA Paper P-1679, September 1982 (SECRET).

5. H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, Princeton, NJ, 1946.

6. D.F. Morrison, "Multivariate Statistical Methods," McGraw-Hill, New York, 1967.

7. H.G. Wolfhard, "Handbook of Aircraft Signatures - Part II Backgrounds," IDA Paper P-1587, July 1982.

8. N.G. Kulgein and H.L. Schick, "Camp Data Analysis," Lockheed Missiles and Space Co., Inc., Report # LMSC-L034162, May 1978 (SECRET).

9. A.J. LaRocca and D.J. Witte, "Handbook of the Statistics of Various Terrain and Water (Ice) Backgrounds from Selected U.S. Locations," Environmental Research Institute of Michigan, Ann Arbor, Michigan, Report # 139900-1-X, January 1980.

10. R.R. Parenti, et al., "The Design of Digital Filters for Resolved Stochastic Targets," Meeting of the IRIS Specialty Group on Targets, Backgrounds, and Discrimination, Infrared Information and Analysis Center, May 1981, p. 393 (SECRET).

11. R. Courant and D. Hilbert, "Methods of Mathematical Physics, V. I.," Interscience, New York, 1953.

12. B.A. Boerschig and D.B. Friedman, "Infrared Local Background Discrimination," Proceedings of the Infrared Information Symposia, v. 2C, No. 1, February 1977.

13. R.A. Steinberg, "Passive Infrared Surveillance: New Methods of Analysis," NRL Memorandum Report # 4078, EOTPO Rep. 55, Electro-Optical Technology Program Office, Naval Research Laboratory, Washington, DC, September 24, 1979.

14. R.A. Steinberg, "Passive Infrared Surveillance Part I: Model Formulation," NRL Report # 8320, EOTPO Rep. 52, September 28, 1979.

15. R.A. Steinberg, "Passive Infrared Surveillance Part II: Threshold Crossing Receivers," NRL Report # 8367, EOTPO Rep. 57, January 24, 1980.

16. R.A. Steinberg, "Infrared Surveillance .1: Statistical Model," Applied Optics, v. 19, January 1, 1980, pp. 77-85.

17. R.A. Steinberg, "Infrared Surveillance .2: Continuous Time Signal Processors," App. Opt., v. 19, May 15, 1980, pp. 1673-1687.

18. R.A. Steinberg, "Quantum-Noise-Limited Signal Processors for Infrared Surveillance," Optical Engineering, v. 20, No. 5, October 1981, pp. 726-735.

19. W.L. Wolfe and G.J. Zissis (Eds.), "The Infrared Handbook," U.S. Government Printing Office, 1978, Ch. 17 (R. Legault, "Reticle and Image Analyses," pp. 17-43--17-46).

Note: In the preparation of this paper, no classified material from the classified references in above listing has been used.

APPENDIX A

OPTIMUM CFAR DISCRIMINATION

APPENDIX A

OPTIMUM CFAR DISCRIMINATION

A. THE GENERAL CASE

It will be assumed that there are N data channels, each providing a radiance measurement proportional to a signal J_i, $i = 1, \ldots, N$. The J_i will be regarded as the components of a vector $\underset{\sim}{J}$ and as coordinates of a point in an N dimensional space.

It will also be assumed that an admissible discrimination process will determine a region R_T in the data space such that all measurement sets representing points in R_T will be regarded as due to a target source and all other measurements as due to clutter. Further, it will be assumed that there is a function $\psi(\underset{\sim}{J})$ and a quantity τ such that the region R_T consists of points $\underset{\sim}{J}$ that satisfy the inequality

$$\psi(\underset{\sim}{J}) < \tau \ . \qquad \text{(A-1)}$$

Suppose that there is a joint probability distribution for the components of $\underset{\sim}{J}$, conditioned on the presence of a target, and an associated probability density $P_T(\underset{\sim}{J})$. Suppose also that the complementary joint probability distribution conditioned on the absence of a target has the density $P_C(\underset{\sim}{J})$. Then the false alarm probability will be given by

$$PFA = \int_{R_T} P_C(\underset{\sim}{J})\, d\underset{\sim}{J} \ , \qquad \text{(A-2)}$$

where the notation is understood to indicate a volume integral over the n dimensional region R_T. Similarly, the probability that a target will be detected if it is present is given by

$$PTD = \int_{R_T} P_T(\underset{\sim}{J})\, d\underset{\sim}{J} \quad . \tag{A-3}$$

For a constant false alarm rate (CFAR) it is necessary to choose the region R_T so that PFA, given by (A-2), is equal to some prescribed constant ϕ. Then the optimum discrimination between targets and clutter will occur when PTD given by (A-3) is maximized subject to the condition that PFA is equal to ϕ.

That is, the problem is to choose the function $\psi(\underset{\sim}{R})$ so as to maximize PTD, the choice being restricted to those functions for which PFA is equal to ϕ. This leads to the variational equation

$$\delta\,[PTD + \lambda\,(\phi - PFA)] = o \quad , \tag{A-4}$$

subject to the condition

$$PFA = \phi \quad , \tag{A-5}$$

where λ is the usual Lagrange multiplier and the variation is taken with respect to $\psi(R)$.

The variation calculated by means of the standard procedure in the calculus of variations after substituting from (A-1), (A-2) and (A-3) leads to the equation

$$\int_{B_T} [P_T(\underset{\sim}{J}) - \lambda\, P_C(\underset{\sim}{J})]\, \delta\psi(\underset{\sim}{J})\, d\underset{\sim}{J} = o \quad , \tag{A-6}$$

where the integral is taken over the hypersurface B_T determined by

$$\psi(\underset{\sim}{J}) = \tau \quad . \tag{A-7}$$

The equation (A-6) must hold for all admissible functions $\delta\psi(\underset{\sim}{J})$; hence, in accordance with the standard argument,

$$P_T(\underset{\sim}{J}) = \lambda \, P_C(\underset{\sim}{J}) \tag{A-8}$$

for all points satisfying (A-7).

The equation (A-8) may be regarded as equivalent to (A-7); hence the relations

$$\begin{gathered} \psi(\underset{\sim}{J}) = \log P_T(\underset{\sim}{J}) - \log P_C(\underset{\sim}{J}) \quad , \\ \tau = \log \lambda \end{gathered} \tag{A-9}$$

provide a satisfactory solution of the variational problem. The constant λ, and therefore τ, can then be determined by solving the equation (A-5) after substituting from (A-2) and (A-9).

B. THE N-VARIATE GAUSSIAN DISTRIBUTION

An important special case, which often has at least an approximate validity, is the case in which $P_T(\underset{\sim}{J})$ and $P_C(\underset{\sim}{J})$ are both N-variate Gaussian probability densities, having the form

$$P(J) = \frac{1}{\sqrt{(2\pi)^n \, |\underset{\approx}{M}|}} \, e^{-\frac{1}{2}(\underset{\sim}{J}-\underset{\sim}{\bar{J}})^t \, \underset{\approx}{M}^{-1} \, (\underset{\sim}{J}-\underset{\sim}{\bar{J}})} \quad , \tag{A-10}$$

where $\bar{\underset{\sim}{J}}$ is the mean vector defined by

$$\bar{\underset{\sim}{J}} = \int_{\infty} \underset{\sim}{J}\, P(\underset{\sim}{J})\, d\underset{\sim}{J} \quad , \tag{A-11}$$

$\underset{\approx}{M}$ is the covariance matrix whose elements M_{ij} are defined by

$$M_{ij} = \int_{\infty} (J_i - \bar{J}_i)\,(J_j - \bar{J}_j)\, P(\underset{\sim}{J})\, d\underset{\sim}{J} \quad , \tag{A-12}$$

$|\underset{\approx}{M}|$ is the determinant of M, and the superscript t in (A-10) indicates the transpose (row) vector. The symbol ∞ in (A-11) and (A-12) means that the integration region for the integrals so labeled is all n space. The densities $P_T(\underset{\sim}{J})$ and $P_C(\underset{\sim}{J})$ are determined completely by their respective means $\bar{\underset{\sim}{J}}_T$ and $\bar{\underset{\sim}{J}}_C$ and their covariance matrices $\underset{\approx}{M}_T$ and $\underset{\approx}{M}_C$ in accordance with the form (A-10).

According to (A-10), (A-9) and (A-1) the optimum CFAR algorithm for declaring a target detection when an N channel measurement set consists of the components of $\underset{\sim}{J}$ is the rule: for a prescribed false alarm probability ϕ a target is present if

$$(\underset{\sim}{J}-\bar{\underset{\sim}{J}}_T)^t\, \underset{\approx}{M}_T^{-1}\, (\underset{\sim}{J}-\bar{\underset{\sim}{J}}_T) - (\underset{\sim}{J}-\bar{\underset{\sim}{J}}_C)^t\, \underset{\approx}{M}_C^{-1}\, (\underset{\sim}{J}-\bar{\underset{\sim}{J}}_C) < \gamma \quad , \tag{A-13}$$

where γ is a constant determined by the condition

$$\frac{1}{\sqrt{(2\pi)^N\, |\underset{\approx}{M}_C|}} \int_{R(\gamma)} e^{-\frac{1}{2}(\underset{\sim}{J}-\bar{\underset{\sim}{J}}_C)^t\, \underset{\approx}{M}_C^{-1}\, (\underset{\sim}{J}-\bar{\underset{\sim}{J}}_C)}\, d\underset{\sim}{J} = \phi \quad . \tag{A-14}$$

The integration region $R(\gamma)$ in (A-14) consists of points $\underset{\sim}{J}$ that satisfy (A-13).

The left side of (A-13) is the difference between two quadratic forms in the quantities $\underset{\sim}{J}-\bar{\underset{\sim}{J}}_T$ and $\underset{\sim}{J}-\bar{\underset{\sim}{J}}_C$. It can be simplified somewhat by a small amount of algebraic manipulation which will reduce it to the sum of a quadratic form in $\underset{\sim}{J}$, a linear form in $\underset{\sim}{J}$ and a constant. In fact, (A-13) can be written

$$\underset{\sim}{J}^t \, \underset{\sim}{\Delta M} \, \underset{\sim}{J} - 2 \, \underset{\sim}{L}^t \, \underset{\sim}{J} < \kappa \quad , \tag{A-15}$$

where ΔM is a matrix given by

$$\underset{\sim}{\Delta M} = \underset{\sim}{M}_T^{-1} - \underset{\sim}{M}_C^{-1} \quad , \tag{A-16}$$

$\underset{\sim}{L}$ is a vector given by

$$\underset{\sim}{L} = \bar{\underset{\sim}{J}}_T \, \underset{\sim}{M}_T^{-1} - \bar{\underset{\sim}{J}}_C \, \underset{\sim}{M}_C^{-1} \quad , \tag{A-17}$$

and κ is a constant given by

$$\kappa = \gamma - \bar{\underset{\sim}{J}}_T^{\,t} \, \underset{\sim}{M}_T^{-1} \, \bar{\underset{\sim}{J}}_T + \bar{\underset{\sim}{J}}_C^{\,t} \, \underset{\sim}{M}_C^{-1} \, \bar{\underset{\sim}{J}}_C \quad .^* \tag{A-18}$$

In deriving (A-15) the fact that the covariance matrices, and therefore their inverses, are symmetric is used.

If $\underset{\sim}{\Delta M}$ is not zero the region in n dimensional space defined by (A-15) is bounded by the n dimensional version of a quadric surface. If n is 2 the boundary is a conic section, i.e., an ellipse, a parabola or a hyperbola.

*If the region R_T is prescribed by the form (A-15) then it is the constant κ that must be determined by (A-14).

If the target and clutter covariance matrices are identical, however, $\Delta\underset{\approx}{M}$ is zero. In that case the region defined by (A-15) is a half space bounded by a hyperplane, i.e., the n dimensional version of a plane.

C. THE UNIVARIATE GAUSSIAN DISTRIBUTION

As an example of how the optimum discrimination algorithm can be formulated in practice it may be useful to consider a special case in detail. The univariate Gaussian distribution is obviously the simplest special case. It is also a useful one to consider because it plays a fundamental role in the construction of optimum linear filters.

When the probability distribution is univariate the mean vector and covariance matrix are actually scalar quantities. Thus, the means associated with the target and clutter distributions are constants $\bar{J}_T$ and $\bar{J}_C$, and variances σ_T^2 and σ_C^2, both of which are also constant, replace the covariance matrices $\underset{\approx}{M}_T$ and $\underset{\approx}{M}_C$.

Then (A-15) becomes

$$\left(\frac{1}{\sigma_T^2} - \frac{1}{\sigma_C^2}\right) J^2 - 2LJ < \kappa \quad , \tag{A-19}$$

where, because of (A-17),

$$L = \frac{\bar{J}_T}{\sigma_T^2} - \frac{\bar{J}_C}{\sigma_C^2} \quad . \tag{A-20}$$

After a substitution from (A-20) the relation (A-19) may be replaced by

$$J^2 - 2\beta J - \gamma \leq 0 \quad , \tag{A-21}$$

where

$$\beta = \frac{\sigma_C^2 \bar{J}_T - \sigma_T^2 \bar{J}_C}{\sigma_C^2 - \sigma_T^2} \qquad \text{(A-22)}$$

and γ is a new constant, replacing κ, to be determined by (A-14).

It follows from (A-21) that

$$\beta - \sqrt{\beta^2+\gamma} < J < \beta + \sqrt{\beta^2+\gamma} \quad . \qquad \text{(A-23)}$$

The interval defined by (A-23) is the one dimensional version of the region that was labeled R_T in the general case and $R(\gamma)$ in the n-variate Gaussian case discussed in Sections A and B of this appendix. That is, (A-23) gives the criterion for declaring that the measurement J is due to a target.

However, before the criterion (A-23) can be used it is still necessary to determine the constant γ. This can be done by using the CFAR condition (A-14), which takes the form

$$\frac{1}{\sigma_C \sqrt{2\pi}} \int_{\beta-\sqrt{\beta^2+\gamma}}^{\beta+\sqrt{\beta^2+\gamma}} e^{-\frac{(J-\bar{J}_C)^2}{2\sigma_C^2}} dJ = \phi \quad ,$$

or, equivalently because of (A-22),

$$\frac{1}{\sqrt{2\pi}} \int_{\alpha-\mu}^{\alpha+\mu} e^{-\frac{1}{2}x^2} dx = \phi \quad , \tag{A-24}$$

where

$$\alpha = \frac{(\bar{J}_T - \bar{J}_C)\ \sigma_C}{\sigma_C^2 - \sigma_T^2} \tag{A-25}$$

and

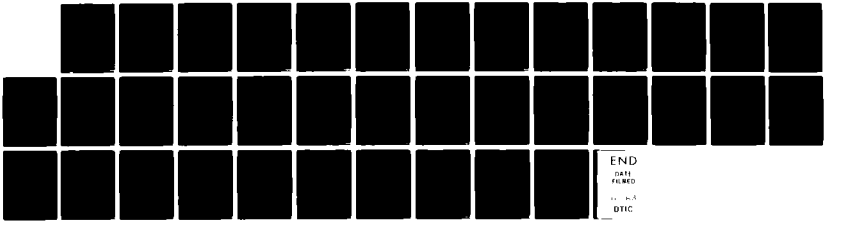
$$\mu = \frac{\sqrt{\beta^2+\gamma}}{\sigma_C} \quad . \tag{A-26}$$

Actually, the simplest procedure now is to determine μ from (A-24) and (A-25) in terms of the CFAR value ϕ. Then, instead of using (A-23), the interval that contains the target declarations can be expressed in terms of μ and β, which is defined in terms of the given probability distribution parameters by (A-22); i.e., a target is declared if

$$\beta - \mu\sigma_C < J < \beta + \mu\sigma_C \quad . \tag{A-27}$$

AD-A128 283 EVALUATION OF INFRARED TARGET DISCRIMINATION ALGORITHMS (U) INSTITUTE FOR DEFENSE ANALYSES ALEXANDRIA VA SCIENCE AND TECHNOLOGY DIV I W KAY APR 83 IDA-P-1714

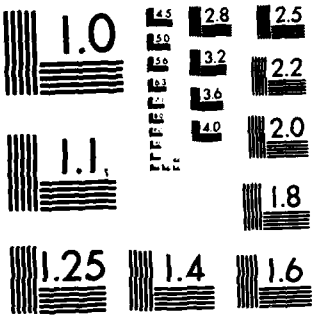
UNCLASSIFIED IDA/HQ-83-25281 F/G 12/1 NL



END

DATE FILMED

DTIC



From the more general definitions (A-3) and (A-10) it follows that

$$PTD = \frac{1}{\sigma_T \sqrt{2\pi}} \int_{\beta-\mu\sigma_C}^{\beta+\mu\sigma_C} e^{-\frac{(J-\bar{J}_T)^2}{2\sigma_T^2}} dJ \quad .$$

With the aid of (A-22) this can be written somewhat more conveniently as

$$PTD = \int_{\hat{\beta}-\hat{\mu}}^{\hat{\beta}+\hat{\mu}} e^{-\frac{x^2}{2}} dx \quad , \qquad \text{(A-28)}$$

where

$$\hat{\mu} = \frac{\mu\sigma_C}{\sigma_T} \qquad \text{(A-29)}$$

and

$$\hat{\beta} = \frac{(\bar{J}_T - \bar{J}_C)\,\sigma_T}{\sigma_C^2 - \sigma_T^2} \quad . \qquad \text{(A-30)}$$

The result of this section can be summarized as follows. Given a required PFA value ϕ, with parameters α defined by (A-25), β defined by (A-22), and μ determined by the equation (A-24), the optimum criterion for a target declaration in the sense of maximizing the PTD is given by (A-27). The corresponding PTD is given by (A-28) in terms of quantities $\hat{\mu}$ defined by (A-29) and $\hat{\beta}$ defined by (A-30).

If $\sigma_T = \sigma_C$ then (A-19) becomes the trivial condition that the interval of possible values of J be divided into two complementary sub-intervals. That is, a constant τ divides the interval

$$-\infty < J < \infty$$

into two intervals

$$-\infty < J < \tau \ , \ \tau < J < \infty \ ,$$

one of which, it will be assumed by the target discrimination rule, contains all values of J, and only those values, that may be attributed to a target source.

If $\Delta\bar{J} > 0$ the value of τ is to be determined by the CFAR condition

$$\phi = \frac{1}{\sigma\sqrt{2\pi}} \int_{\tau}^{\infty} e^{-\frac{(J-\bar{J}_C)^2}{2\sigma^2}} dJ = \frac{1}{\sqrt{2\pi}} \int_{\frac{\tau-\bar{J}_C}{\sigma}}^{\infty} e^{-\frac{x^2}{2}} dx \ ,$$

where σ is the common value of σ_C and σ_T. That is, for ν such that

$$\frac{1}{\sqrt{2\pi}} \int_{\nu}^{\infty} e^{-\frac{x^2}{2}} dx = \phi \qquad \text{(A-31)}$$

τ will be determined by

$$\tau = \sigma\nu + \bar{J}_C \quad . \qquad \text{(A-32)}$$

Then a target is declared whenever

$$J > \tau \quad .$$

If $\Delta\bar{J} < 0$ then ν is still defined by (A-31) but (A-32) must be replaced by

$$\tau = -\sigma\nu + \bar{J}_C \quad . \qquad \text{(A-33)}$$

Then a target is declared if

$$J < \tau \quad .$$

In either case the PTD can be calculated by integrating

$$P_T(J) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(J-\bar{J}_T)^2}{2\sigma^2}} \quad .$$

over the interval in which targets are declared. The result is given by

$$PTD = \frac{1}{\sqrt{2\pi}} \int_{\nu - \frac{|\Delta \bar{J}|}{\sigma}}^{\infty} e^{-\frac{x^2}{2}} dx \, . \qquad (A-34)$$

APPENDIX B

POWER SPECTRAL DENSITY CALCULATIONS

APPENDIX B

POWER SPECTRAL DENSITY CALCULATIONS

Chapter II discusses two correlation functions that are sometimes suggested as models for the spatial distribution of background clutter. One is

$$K_{x,y}\ (\underset{\sim}{r},\underset{\sim}{r}') = \exp\left(-\frac{|x-x'|}{L_x} - \frac{|y-y'|}{L_y}\right) , \qquad \text{(B-1)}$$

which is anisotropic even when $L_x = L_y$. The other is

$$K(\underset{\sim}{r},\underset{\sim}{r}') = \exp\left(-\frac{|\underset{\sim}{r}-\underset{\sim}{r}'|}{L}\right) , \qquad \text{(B-2)}$$

which is isotropic.

The power spectral density (or Wiener spectrum) for $K_{x,y}\ (\underset{\sim}{r},\underset{\sim}{r}')$ is given by the Fourier transform

$$W_{x,y}\ (\underset{\sim}{k}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{|u|}{L_x} - \frac{|v|}{L_y} + i\underset{\sim}{k}^t\ \underset{\sim}{u}\right) dudv, \qquad \text{(B-3)}$$

where $\underset{\sim}{k}$ is the wave number vector with components (k_x, k_y) and $\underset{\sim}{u}$ is the displacement vector with components

$$u = x-x',\ v = y-y' \ .$$

Since

$$\underset{\sim}{k}^t \underset{\sim}{u} = k_x u + k_y v \quad ,$$

the double integral in (B-3) is a product of two single integrals that are easy to evaluate individually. The result of the evaluation is

$$W_{x,y}(\underset{\sim}{k}) = \left(\frac{L_x}{1-ik_xL_x} + \frac{L_x}{1+ik_xL_x}\right)\left(\frac{L_y}{1-ik_yL_y} + \frac{L_y}{1+ik_yL_y}\right) \tag{B-4}$$

$$= \frac{4L_xL_y}{(1+k_x^2L_x^2)(1+k_y^2L_y^2)} \quad .$$

The power spectral density for $K(\underset{\sim}{r},\underset{\sim}{r}')$ is given by the Fourier transform

$$W(\underset{\sim}{k}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{|\underset{\sim}{u}|}{L} + i\underset{\sim}{k}^t\underset{\sim}{u}\right) dudv \quad , \tag{B-5}$$

which, after changing variables to polar coordinates (P, o), becomes

$$W(\underset{\sim}{k}) = \int_0^{\infty} e^{-\frac{\rho}{L}} \int_0^{2\pi} e^{ik\rho\cos(\theta-\phi)}\, d\theta\rho d\rho \quad , \tag{B-6}$$

where (k, ϕ) are the polar coordinates of the vector $\underset{\sim}{k}$. The inner, angular, integral is independent of ϕ because the

integration interval is exactly one period of the integrand which, of course, is periodic in θ; therefore, ϕ can be set equal to zero. The integral over θ can then be recognized as a well known representation for the Bessel function of order zero. Thus, (B-6) may be written

$$W(k) = 2\pi \int_0^\infty e^{-\frac{\rho}{L}} J_0(k\rho)\, \rho d\rho \quad .$$

With the aid of a standard table of integrals (e.g., Ref. B-1, p. 712) this can be recognized as equivalent to

$$W(k) = \frac{2\pi L^2}{(1+k^2L^2)^{3/2}} \quad . \qquad \text{(B-7)}$$

As Ref. B-2 points out, if the background distribution appears to be isotropic when viewed at a 90 deg depression angle (i.e., vertically), when it is viewed at an angle that is less than 90 deg it should appear to be anisotropic. This is because there will be an apparent change of scale in the cross-track but not in the in-track direction. In such a case the natural generalization of the correlation function $K(\underset{\sim}{r},\underset{\sim}{r}')$ given by (B-2) is a function of the form

$$K_y(\underset{\sim}{r},\underset{\sim}{r}') = \exp\left[-\frac{1}{L}\sqrt{(x-x')^2+\alpha^2(y-y')^2}\right] \quad . \qquad \text{(B-8)}$$

The corresponding power spectral density will be given by

$$W_y(\underset{\sim}{k}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[i\underset{\sim}{k}^t\underset{\sim}{u} - \frac{1}{L}\sqrt{u^2 + \alpha^2 v^2}\right] dudv \quad ,$$

which, after a change of variable, becomes

$$W_y(\underset{\sim}{k}) = \frac{1}{\alpha}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(i\hat{\underset{\sim}{k}}^t\underset{\sim}{u} - \frac{|\hat{\underset{\sim}{u}}|}{L}\right) d\hat{u}d\hat{v} \quad , \qquad \text{(B-9)}$$

where $\hat{\underset{\sim}{k}}$ is the vector with components $\left(k_x, \frac{k_y}{\alpha}\right)$ and $\hat{\underset{\sim}{u}}$ is the vector with components $(u, \alpha v)$. Since (B-9) has exactly the form of (B-5) its value can be obtained by inspection of (B-7). Thus,

$$W_y(\underset{\sim}{k}) = \frac{2\pi\alpha^2L^2}{(\alpha^2 + \alpha^2L^2k_x^2 + L^2k_y^2)^{3/2}} \quad . \qquad \text{(B-10)}$$

Isotropic background distributions with Wiener spectra for which the functional forms differ from that in (B-7) apparently occur more frequently than not. A natural generalization of the Wiener spectrum given by (B-7) that produces an infinite class of possible correlation functions can be obtained by replacing the exponent $\frac{3}{2}$ in the denominator on the right side with any positive number ν. The corresponding correlation function would then be given by

$$K(\underset{\sim}{r},\underset{\sim}{r}';\ \nu) = L^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp\left[-i\underset{\sim}{k}^t\ (\underset{\sim}{r}-\underset{\sim}{r}')\right]}{(1+k^2L^2)^{\nu}}\ dk_x\ dk_y$$

(B-11)

$$= L^2 \int_0^{\infty} \frac{J_o(k|\underset{\sim}{r}-\underset{\sim}{r}'|)}{(1+k^2L^2)^{\nu}}\ kdk \quad .$$

The integral in (B-11) can be evaluated with the aid of a formula on p. 488 of Ref. B-3. The result is

$$K(\underset{\sim}{r},\underset{\sim}{r}';\nu) = \frac{1}{\Gamma(\nu)} \left(\frac{|\underset{\sim}{r}-\underset{\sim}{r}'|}{2L}\right)^{\nu-1} K_{\nu-1}\left(\frac{|\underset{\sim}{r}-\underset{\sim}{r}'|}{L}\right) \qquad \text{(B-12)}$$

where the function on the right having the form $K_n(x)$ is a modified Bessel function. The modified Bessel function $K_n(x)$ can be expressed in terms of elementary functions when the order n is an odd multiple of $\frac{1}{2}$ (cf. Ref. B-3 , p. 444); e.g.,

$$K_{-\frac{1}{2}}(x) = K_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}}\ e^{-x} \quad ,$$

$$K_{-\frac{3}{2}}(x) = K_{\frac{3}{2}}(x) = \sqrt{\frac{\pi}{2x}}\ e^{-x}\ (1 + \tfrac{1}{x}) \quad .$$

APPENDIX B

REFERENCES

B-1. I.S. Gradshteyn and I.M. Ryzlik, "Table of Integrals, Series, and Products (4th Ed.)," Academic Press, New York, 1965.

B-2. R.R. Parenti, et al., "The Design of Digital Filters for Resolved Stochastic Targets," Meeting of the IRIS Specialty Group on Targets, Backgrounds, and Discrimination," Infrared Information and Analysis Center, May 19[illegible] p. 393 (SECRET).

B-3. M. Abromowitz and I.A. Stegun (Eds.), "Handbook of Mathematical Functions," National Bureau of Standards Series 55, U.S. Government Printing Office, June 1964 (10th Printing, December 1972).

Note: In the preparation of this paper, no classified material from the classified references in above listing has been used.

APPENDIX C

CANONICAL VARIABLES FOR SPATIAL CHANNELS

APPENDIX C

CANONICAL VARIABLES FOR SPATIAL CHANNELS

The mathematical model for spatial channels that was introduced in Chapter II assumes a target covariance matrix $\underset{\sim}{M}_T$ given by

$$\underset{\sim}{M}_T = \underset{\sim}{M}_c - \Delta\underset{\sim}{M} \quad , \tag{C-1}$$

where $\underset{\sim}{M}_c$ has the elements M_{ij} and $\Delta\underset{\sim}{M}$ has the elements Δ_{ij} given by

$$\Delta_{ij} = M_{jo}\,\delta_{io} + M_{io}\,\delta_{jo} - M_{oo}\,\delta_{io}\,\delta_{jo} - \sigma_T^2\,\delta_{io}\,\delta_{jo} \; . \tag{C-2}$$

In (C-2) and throughout this Appendix the subscripts are understood to be two component vectors, and quantities δ_{ik} are Kronecker deltas which are equal to one if the subscript vectors i and k are identical in both of their components but are otherwise zero. A zero subscript represents the zero vector, both of whose components are zero. Any sum that is indicated over a subscript will mean that a double sum is to be taken independently over both components of the subscript vector.

In the case of an n by n pixel window there will be n^2 channels, one for every pixel. Each of the subscript vector components independently takes on n values, so that the vector, itself, takes on n^2 values, one for every channel.

As observed in Chapter III, for optimum CFAR discrimination in general it is sometimes useful to transform the variables

associated with the natural channels to new variables in terms of which the target and clutter covariance matrices are both diagonal. This can be done by solving the eigenvalue problem*

$$(\underset{\approx}{M}_c - \lambda_m \underset{\approx}{M}_T)\, \underset{\sim}{Y}_m = 0 \quad . \tag{C-3}$$

The purpose of this Appendix is to derive algorithms for calculating the n^2 eigenvalues λ_m and eigenvectors $\underset{\sim}{Y}_m$ for the spatial channel model defined by (C-1) and (C-2) in the case of an n by n pixel window.

First of all, it is evident from (C-1) and (C-3) that any set of linearly independent vectors $\underset{\sim}{Y}_m$ that satisfy

$$\Delta\underset{\approx}{M}\, \underset{\sim}{Y}_m = 0 \tag{C-4}$$

will be a set of distinct eigenvectors asscciated with the common eigenvalue

$$\lambda_m = 1 \quad .$$

In fact, by using (C-2) explicitly in (C-4) one finds that any vector with components Y_k that satisfy the equations

$$\begin{aligned} Y_o &= 0 \ , \\ \sum_k M_{ko}\, Y_k &= 0 \end{aligned} \tag{C-5}$$

will be such an eigenvector. For an n by n pixel window the sum in (C-5) contains at most n^2-1 non-zero terms since the term corresponding to k = o vanishes.

*Cf. Ref. (C-1), pp. 37-41.

If it is assumed that the clutter probability distribution is non-degenerate the covariance matrix $\underset{\sim}{\underset{\sim}{M}}_C$ is non-singular. Then the n^2 column vectors of $\underset{\sim}{\underset{\sim}{M}}_C$ are linearly independent.

From the column vectors $\underset{\sim}{U}_i$, $i \neq o$, whose components are M_{ki}, if there are not already n^2-2 values of i for which M_{oi} is zero, it is possible to form n^2-2 new linearly independent vectors $\underset{\sim}{V}_i$ by defining the components of $\underset{\sim}{V}_i$ by

$$V_{ki} = M_{ki} - \frac{M_{ki}}{M_{oj}} M_{kj}, \quad i \neq o, \quad i \neq j \quad , \qquad (C-6)$$

where j is any fixed subscript vector such that M_{oj} is not zero. The vectors $\underset{\sim}{V}_i$ defined by (C-6) all have the component V_{oi} equal to zero. Together with the vector $\underset{\sim}{V}_o$ whose components are M_{ko} for $k \neq o$ and zero in place of M_{co}, the vectors $\underset{\sim}{V}_i$ form a set of n^2-1 linearly independent vectors, all having zero for the component labeled with the subscript o.

Applying the Gram-Schmidt orthogonalization process* to the n^2-1 vectors $\underset{\sim}{V}_i$ leads to a set of n^2-2 orthogonal vectors $\underset{\sim}{Y}_i$, each of whose components provide a different solution of the equations (C-5). These vectors $\underset{\sim}{Y}_i$ are therefore n^2-2 linearly independent eigenvectors of (C-3), all corresponding to the same eigenvalue, one.

The Gram-Schmidt process is equivalent to a recursive algorithm that is computationally efficient and easy to implement. Given a set of N linearly independent vectors $\underset{\sim}{X}_n$ and a defined inner product $(\underset{\sim}{U}, \underset{\sim}{V})$ for any pair of vectors $\underset{\sim}{U}$ and $\underset{\sim}{V}$, the following recursion relation generates a set of N vectors $\underset{\sim}{Y}_n$ that are mutually orthogonal with respect to the inner product and such that $\underset{\sim}{Y}_o$ is equal to the vector identified as $\underset{\sim}{X}_o$ in the original set:

---

*Cf. Ref. (C-2), p. 230.

$$\underset{\sim}{Y}_o = \underset{\sim}{Z}_o = \underset{\sim}{X}_o \quad ,$$

$$\underset{\sim}{Z}_m = \underset{\sim}{X}_m - \sum_{\nu=o}^{m-1} (\underset{\sim}{X}_m, \underset{\sim}{Y}_\nu)\, \underset{\sim}{Y}_\nu, \quad m=1, \ldots, N \qquad (C\text{-}7)$$

$$Y_m = \frac{Z_m}{\sqrt{(Z_m, Z_m)}} \quad , \quad m=1, \ldots, N-1 \; .$$

To calculate the eigenvectors of (C-3) it is only necessary to identify $\underset{\sim}{X}_o$ with the vector whose components are defined to be M_{ko}, to identify the number N with n^2-1, and to define the inner product as the usual scalar product of two vectors; i.e., in terms of column vectors $\underset{\sim}{C}$ and $\underset{\sim}{K}$ the inner product will be defined by

$$(C, K) = \underset{\sim}{C}^t \underset{\sim}{K} = \sum_{\nu=o}^{N-1} C_\nu K_\nu \quad , \qquad (C\text{-}8)$$

where the superscript t means "transpose".

The eigenvalue equation (C-3) has two additional eigenvalues for λ_m that are not equal to one, with corresponding eigenvectors $\underset{\sim}{U}$ and $\underset{\sim}{V}$. Together with the n^2-2 vectors $\underset{\sim}{Y}_m$, $m \neq o$, $\underset{\sim}{U}$ and $\underset{\sim}{V}$ complete a set of n^2 linearly independent eigenvectors that satisfy (C-3) when the corresponding eigenvalues are used for λ_m.

To deal with the problem of finding the two new eigenvalues and eigenvectors it is convenient to define a new inner product by

$$(\underset{\sim}{C}, \underset{\sim}{K}) = \underset{\sim}{C}^t \underset{\approx}{M} \underset{\sim}{K} \quad , \qquad (C\text{-}9)$$

which is a bilinear form relative to the covariance matrix $\underset{\approx}{M}_C$. Since $\underset{\approx}{M}_C$ must be positive definite the inner product defined by (C-9) has all of the properties that are necessary for the inner product operation. In particular, it follows from the standard argument* that, relative to the new inner product, $\underset{\sim}{U}$ and $\underset{\sim}{V}$ will each be orthogonal to all of the eigenvectors $\underset{\sim}{Y}_m$ and to each other, i.e.,

$$(\underset{\sim}{U}, \underset{\sim}{V}) = (\underset{\sim}{U}, \underset{\sim}{Y}_m) = (\underset{\sim}{V}, \underset{\sim}{Y}_m) = 0, \; m \neq 0 \quad . \qquad \text{(C-10)}$$

The orthogonality of two different eigenvectors in the sense of (C-10) in terms of the inner product defined by (C-9) depends upon the corresponding eigenvalues being different. Therefore, it does not necessarily hold among the first n^2-2 eigenvectors $\underset{\sim}{Y}_m$.

The orthogonality property (C-10) can be used to find $\underset{\sim}{U}$ and $\underset{\sim}{V}$ and the corresponding eigenvalues. First, it is necessary to define two vectors $\underset{\sim}{P}$ and $\underset{\sim}{Q}$ which are orthogonal to the $\underset{\sim}{Y}_m$ and to each other. Then $\underset{\sim}{P}$, $\underset{\sim}{Q}$ and the $\underset{\sim}{Y}_m$ form a set of n^2 linearly independent vectors which span the n^2 dimensional vector space.

It will be found that suitable candidates for $\underset{\sim}{P}$ and $\underset{\sim}{Q}$ are the vectors whose components are given by

$$\begin{aligned} P_i &= \delta_{io} \quad , \\ Q_i &= M_{io} - S\,\delta_{io} - \sum_{\nu k l} \left(M_{ko}\, M_{kl}\, Y_{l\nu} \right) Y_{i\nu} \quad , \end{aligned} \qquad \text{(C-11)}$$

where

$$S = \frac{1}{M_{oo}} \sum_k M_{ko}^2$$

*Cf. Ref. (C-1), pp. 37-41.

and the $\underset{\sim}{V}_{i\nu}$ are the components of the vectors $\underset{\sim}{V}_{\nu}$. To verify that $\underset{\sim}{P}$ and $\underset{\sim}{Q}$ whose components are defined by (C-11) satisfy the orthogonality conditions (C-10) when substituted for $\underset{\sim}{U}$ and $\underset{\sim}{V}$ it is only necessary to substitute from (C-11) into (C-10) and make use of (C-5) and the symmetry of the covariance matrix $\underset{\approx}{M}_C$.

It follows from the orthogonality property (C-10) of the eigenvectors that $\underset{\sim}{U}$ and $\underset{\sim}{V}$ must belong to the subspace spanned by $\underset{\sim}{P}$ and $\underset{\sim}{Q}$. That is, if $\underset{\sim}{P}$ is not already an eigenvector, as is usually the case, then either eigenvector can be written in the form

$$\underset{\sim}{U} = \alpha \underset{\sim}{P} + \underset{\sim}{Q} \quad , \tag{C-12}$$

where α is a scalar constant to be determined.

Because of (C-1) the eigenvalue equation (C-3) can be written

$$(1-\lambda)\, \underset{\approx}{M}_C \underset{\sim}{Y} + \lambda \Delta \underset{\approx}{M} \underset{\sim}{Y} = 0 \quad . \tag{C-13}$$

If $\underset{\sim}{U}$ given by (C-12) is substituted for $\underset{\sim}{Y}$ in (C-13) and the resulting equation is multiplied on the left by $\underset{\sim}{P}^t$ or by $\underset{\sim}{Q}^t$ the first or the second of the two scalar equations

$$\begin{aligned} (1-\lambda)\, (\alpha \Pi_{11} + \Pi_{12}) + \lambda(\alpha \Gamma_{11} + \Gamma_{12}) &= 0 \quad , \\ (1-\lambda)\, (\alpha \Pi_{12} + \Pi_{22}) + \lambda(\alpha \Gamma_{12} + \Gamma_{22}) &= 0 \end{aligned} \tag{C-14}$$

results, where

$$\begin{aligned} \Pi_{11} &= \underset{\sim}{P}^t \underset{\approx}{M}_C \underset{\sim}{P}, \quad \Pi_{12} = \underset{\sim}{P}^t \underset{\approx}{M}_C \underset{\sim}{Q}, \quad \Pi_{22} = \underset{\sim}{Q}^t \underset{\approx}{M}_C \underset{\sim}{Q} \quad , \\ \Gamma_{11} &= \underset{\sim}{P}^t \Delta \underset{\approx}{M} \underset{\sim}{P}, \quad \Gamma_{12} = \underset{\sim}{P}^t \Delta \underset{\approx}{M} \underset{\sim}{Q}, \quad \Gamma_{22} = \underset{\sim}{Q}^t \Delta \underset{\approx}{M} \underset{\sim}{Q} \quad . \end{aligned} \tag{C-15}$$

By eliminating α from (C-14), using

$$\alpha = \frac{(\lambda-1)\,(\Pi_{12}^2 - \Pi_{11}\,\Pi_{22}) + \lambda\,(\Pi_{11}\,\Gamma_{22} - \Pi_{12}\,\Gamma_{12})}{\lambda(\Pi_{12}\,\Gamma_{11} - \Pi_{11}\,\Gamma_{12})} \qquad \text{(C-16)}$$

obtained from the second equation to substitute into the first, a quadratic equation in λ,

$$a\,\lambda^2 + b\,\lambda + c = 0 \quad , \qquad \text{(C-17)}$$

results. After some tedious but straightforward algebra it will be found that the coefficients of (C-17) are given by

$$\begin{aligned} a &= \Pi_{11}\,(\Pi_{11}\,\Pi_{22} - \Pi_{12}^2 + \Gamma_{11}\,\Gamma_{22} - \Gamma_{12}^2 + 2\,\Pi_{12}\,\Gamma_{12} - \Pi_{11}\,\Gamma_{22} - \Gamma_{11}\,\Pi_{22}) \quad , \\ b &= \Pi_{11}\,(\Pi_{11}\,\Gamma_{22} + \Gamma_{11}\,\Pi_{22} - 2\,\Pi_{12}\,\Gamma_{12} + 2\,\Pi_{12}^2 - 2\,\Pi_{11}\,\Pi_{22}) \quad , \qquad \text{(C-18)} \\ c &= \Pi_{11}\,(\Pi_{11}\,\Pi_{22} - \Pi_{12}^2) \quad . \end{aligned}$$

In the definition of a, b, c by (C-18) the common factor Π_{11} can be omitted since it has no effect on the solution of (C-17).

With the aid of (C-2) and (C-11) the Π_{ij} and T_{ij} can be calculated explicitly from (C-15). The results are

$$\begin{aligned} &\Pi_{11} = M_{oo},\ \Pi_{12} = 0,\ \Pi_{22} = \sum_{ij} M_{io}\,M_{ij}\,M_{jo} - S^2\,M_{oo} \quad , \\ &\Gamma_{11} = M_{oo} - \sigma_T^2,\ \Gamma_{12} = \sigma_T^2\,(S - M_{oo}),\ \Gamma_{22} = -\,(M_{oo}+\sigma_T^2)\,(S-M_{oo})^2 \quad , \end{aligned} \qquad \text{(C-19)}$$

where S is given by the last equation in (C-11).

It is of some interest to obtain, explicitly, the discriminant d of the quadratic equation (C-17) after removal of the common factor II_{11}. The result, calculated from (C-18), is

$$d = b^2 - 4ac = (II_{11}\,\Gamma_{22} + \Gamma_{11}\,II_{22})^2 + 4\,II_{11}\,II_{22}\,(\Gamma_{12}^2 - \Gamma_{11}\,\Gamma_{22})\,. \qquad \text{(C-20)}$$

From the definitions (C-15) and the fact that $\underset{\approx}{M}_C$ is a positive definite matrix it follows that II_{11} and II_{22} are both positive. Thus, if Γ_{12} on the right side of (C-20) is replaced by zero the effect will be to decrease the right side of the equation. That is,

$$d \geq (II_{11}\,\Gamma_{22} + \Gamma_{11}\,II_{22})^2 - 4\,II_{11}\,II_{22}\,\Gamma_{11}\,\Gamma_{22} = (II_{11}\,\Gamma_{22} - \Gamma_{11}\,II_{22})^2 \geq 0. \qquad \text{(C-21)}$$

In other words, according to (C-21) the discriminant is always non-negative. Therefore, the quadratic equation (C-17) for the eigenvalues λ has only real roots, which is certainly a requirement. In fact, a fortiori, since $\underset{\approx}{M}_C$ and $\underset{\approx}{M}_T$ are both positive definite the equation (C-3) can only be satisfied for positive real values of λ.

A reference to the form of (C-20) and of (C-21) indicates that the discriminant vanishes; i.e., the roots of (C-17) will be equal, only if Γ_{12} is zero and $II_{11}\,\Gamma_{22} = \Gamma_{11}\,II_{22}$, or if II_{22} and Γ_{22} are both zero. A reference to (C-19) confirms that the first pair of conditions either imply the second pair, which are satisfied only when

$$S = M_{oo} \qquad \text{(C-22)}$$

and

$$\sum_{ij} M_{io}\,M_{ij}\,M_{jo} = M_{oo}^3 \quad ,$$

or else they imply that

$$S = M_{oo} \tag{C-23}$$

and

$$M_{oo} = \sigma_T^2 \quad .$$

Thus, the eigenvalues will be equal if and only if (C-22) or (C-23) is satisfied.

If the eigenvalues are equal the corresponding independent eigenvectors are $\underset{\sim}{P}$ and $\underset{\sim}{Q}$ defined by (C-11). If the eigenvalues are not equal their corresponding eigenvectors are given by (C-12) with the respective values of α given by (C-16).

The n^2-2 linearly independent vectors $\underset{\sim}{Y}_i$ that were obtained from the $\underset{\sim}{V}_i$ defined by (C-6) are eigenvectors satisfying the equation (C-3), all corresponding to the eigenvalue $\lambda_m = 1$. They are also mutually orthogonal with respect to the usual inner product defined by (C-8).

The set of all eigenvector solutions of (C-3) consists of the $\underset{\sim}{Y}_i$ and the two additional vectors $\underset{\sim}{U}$ and $\underset{\sim}{V}$ that are linear combinations of $\underset{\sim}{P}$ and $\underset{\sim}{Q}$, whose components are defined by (C-11). However, in order to use these eigenvectors to construct a principal axis transformation that simultaneously diagonalizes $\underset{\approx}{M}_C$ and $\underset{\approx}{M}_T$, which was the original purpose of the analysis in this appendix, a further step is necessary. The eigenvectors must be mutually orthogonal with respect to the inner product defined by (C-9).

This condition is satisfied by $\underset{\sim}{U}$, $\underset{\sim}{V}$ and any one of the $\underset{\sim}{Y}_i$, because they correspond to different eigenvalues, but it is not necessarily satisfied by the $\underset{\sim}{Y}_i$. However, n^2-2 linear combinations of the $\underset{\sim}{Y}_i$ can be found that are mutually orthogonal with respect to the inner product defined by (C-9).

The Gram-Schmidt orthogonalization process, e.g., in the form of the algorithm given by the recursion relations (C-7), will accomplish this objective when it is applied to the $\underset{\sim}{Y}_i$, using (C-9) instead of (C-8).

The vectors $\underset{\sim}{U}$, $\underset{\sim}{V}$ and the resulting linear combinations of the $\underset{\sim}{Y}_i$ are then the column vectors of a matrix $\underset{\approx}{T}$ which provides the desired principal axis transformation. That is,

$$\underset{\approx}{T}^t \underset{\approx}{M}_C \underset{\approx}{T} \text{ and } \underset{\approx}{T}^t \underset{\approx}{M}_T \underset{\approx}{T}$$

will both be diagonal matrices, as required.

REFERENCES

C-1. R. Courant and D. Hilbert, "Methods of Mathematical Physics, V.I.," Interscience, New York, 1953.

C-2. K. Hoffman and R. Kunze, "Linear Algebra," Prentice-Hall, Englewood Cliffs, N.J., 1961.

APPENDIX D

RATIOS OF MULTI-VARIATE GAUSSIAN DISTRIBUTED VARIABLES

APPENDIX D

RATIOS OF MULTI-VARIATE GAUSSIAN DISTRIBUTED VARIABLES

1. Probability Density Functions

Reference D-1 contains a derivation of the joint probability density of the ratios

$$X_1 = \frac{J_1}{J_2} \ , \ X_2 = \frac{J_3}{J_2}$$

when J_1, J_2, J_3 have a tri-variate Gaussian distribution. In this appendix the derivation will be generalized to cover the case of variables $J_1,\ldots,J_N$ with an N-variate Gaussian distribution.

That is, it will be assumed that there is a joint probability density function given by

$$P(\underset{\sim}{J}) = \frac{1}{(2\pi)^{N/2}|\underset{\approx}{M}|} \, e^{-\frac{1}{2}(\underset{\sim}{J}-\bar{\underset{\sim}{J}})^t \underset{\approx}{M}^{-1}(\underset{\sim}{J}-\bar{\underset{\sim}{J}})} \ , \qquad \text{(D-1)}$$

where $\underset{\sim}{J}$ is an N-dimensional vector, $\bar{\underset{\sim}{J}}$ is the N-dimensional mean vector, $\underset{\approx}{M}$ is the N by N covariance matrix relative to the probability distribution for $\underset{\sim}{J}$ and $|\underset{\approx}{M}|$ is the determinant of $\underset{\approx}{M}$. The problem is to determine the joint probability density function $P_R(\underset{\sim}{X})$ for the ratios

$$X_i = \frac{J_i}{J_N} \ , \ i=1,\ldots,N-1 \ ,$$

which are components of an (N-1)-dimensional vector $\underset{\sim}{X}$.

The argument used in Ref. B-1 for the case of three variables can be extended to cover the general case of N variables. The first step is to define the change of variables

$$\begin{aligned} J_i &= UX_i, \quad i=1,\ldots,N-1 \quad , \\ J_N &= U \quad . \end{aligned} \tag{D-2}$$

The Jacobean for the transformation (D-2) is then given by

$$\frac{\partial (J_1,\ldots,J_N)}{\partial (X_1,\ldots,X_{N-1},U)} = \begin{vmatrix} U,0,0,\ldots,0,X_1 \\ 0,U,0,\ldots,0,X_2 \\ \cdot\ \cdot\ \cdot\ \ldots\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot\ \ldots\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot\ \ldots\ \cdot\ \cdot \\ 0,0,\ldots\ 0,1 \end{vmatrix} = U^{N-1} \quad . \tag{D-3}$$

The (N-1)-variate probability density function for the ratios X_i is given by

$$P_R(\underset{\sim}{X}) = \int_{-\infty}^{\infty} P(\underset{\sim}{J}) \left| \frac{\partial (J_1,\ldots,J_N)}{\partial (X_1,\ldots,X_{N-1},U)} \right| dU \quad . \tag{D-4}$$

Substitutions from (D-1), (D-2), and (D-3) into (D-4) then lead to the result

$$P_R(\underset{\sim}{X}) = \frac{1}{(2\pi)^{N/2} \sqrt{|\underset{\approx}{M}|}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} Q(X_i,\bar{J}_i,U)} |U|^{N-1} \, dU \quad , \tag{D-5}$$

where

$$Q(X_i,\bar{J}_i,U) = \sum_{i,j=1}^{N} A_{ij}\ (UY_i-\bar{J}_i)\ (UY_j-\bar{J}_j) \quad ,$$

in which the coefficients A_{ij} are the elements of the inverse covariance matrix $\underset{\approx}{M}^{-1}$ and, by definition,

$$\begin{aligned} Y_i &= X_i \ , \ i=1,\ldots,N-1, \\ Y_N &= 1 \ . \end{aligned} \tag{D-6}$$

Then a straightforward calculation provides the result

$$P_R(\underset{\approx}{X}) = \frac{e^{-\frac{1}{2}C}}{(2\pi)^{N/2}\sqrt{|\underset{\approx}{M}|}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}AU^2+BU} |U|^{N-1}\, dU \quad , \tag{D-7}$$

where

$$A = \underset{\sim}{Y}^t\ \underset{\approx}{M}^{-1}\ \underset{\sim}{Y},\ B = \bar{\underset{\sim}{J}}^t\ \underset{\approx}{M}^{-1}\ \underset{\sim}{Y},\ C = \bar{\underset{\sim}{J}}^t\ \underset{\approx}{M}^{-1}\ \bar{\underset{\sim}{J}} \quad ,$$

in which $\underset{\sim}{Y}$ is the N-dimensional vector whose components are given by (D-6).

When N is an odd number (D-7) can be written

$$P_R(\underset{\approx}{X}) = \frac{e^{-\frac{1}{2}C}}{(2\pi)^{N/2}\sqrt{|\underset{\approx}{M}|}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}AU^2+BU} U^{N-1}\, dU \quad . \tag{D-8}$$

However, when N is even

$$P_R(\underset{\sim}{X}) = \frac{e^{-\frac{1}{2}C}}{(2\pi)^{N/2}\sqrt{|\underset{\approx}{M}|}}\left[\int_0^\infty e^{-\frac{1}{2}AU^2+BU} U^{N-1}\,dU - \int_{-\infty}^0 e^{-\frac{1}{2}AU^2+BU} U^{N-1}dU\right]$$

$$= \frac{e^{-\frac{1}{2}C}}{(2\pi)^{N/2}\sqrt{|\underset{\approx}{M}|}}\left[\int_0^\infty e^{-\frac{1}{2}AU^2+BU} U^{N-1}\,dU + \int_0^\infty e^{-\frac{1}{2}AU^2-BU} U^{N-1}\,dU\right]$$

(D-9)

$$= \frac{e^{\frac{B^2-AC}{2A}}}{(2\pi)^{N/2}\sqrt{|\underset{\approx}{M}|}}\left[\int_0^\infty e^{-\frac{A}{2}\left(U-\frac{B}{A}\right)^2} U^{N-1}\,dU + \int_0^\infty e^{-\frac{A}{2}\left(U+\frac{B}{A}\right)^2} U^{N-1}\,dU\right]$$

$$= \frac{e^{\frac{B^2-AC}{2A}}}{(2\pi A)^{N/2}\sqrt{|\underset{\approx}{M}|}}\left[\int_{-\frac{B}{\sqrt{A}}}^\infty e^{-\frac{1}{2}W^2}\left(W+\frac{B}{\sqrt{A}}\right)^{N-1} dW + \int_{\frac{B}{\sqrt{A}}}^\infty e^{-\frac{1}{2}W^2}\left(W-\frac{B}{\sqrt{A}}\right)^{N-1} dW\right].$$

Similar steps taken with (D-8), for the case in which N is odd, lead to the result

$$P_R(\underset{\approx}{X}) = \frac{e^{\frac{B^2-AC}{2A}}}{(2\pi A)^{N/2}\sqrt{|\underset{\approx}{M}|}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}W^2} \left(W + \frac{B}{\sqrt{A}}\right)^{N-1} dW \,. \qquad \text{(D-10)}$$

With the aid of the binomial expansion theorem (D-10) can be written

$$P_R(\underset{\approx}{X}) = \frac{e^{\frac{B^2-AC}{2A}}}{(2\pi A)^{N/2}\sqrt{|\underset{\approx}{M}|}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}W^2} \sum_{r=0}^{N-1} \binom{N-1}{r} B^{N-1-r} A^{\frac{1}{2}(r+1-N)} W^r \, dW \qquad \text{(D-11)}$$

$$= \frac{e^{\frac{B^2-AC}{2A}}}{(2\pi)^{\frac{N-1}{2}}\sqrt{|\underset{\approx}{M}| A^N}} \sum_{r=0}^{N-1} \binom{N-1}{r} B^{N-1-r} A^{\frac{1}{2}(r+1-N)} \mu_r \,,$$

where μ_r is the r^{th} moment of the standard normal probability distribution, i.e., a Gaussian distribution with zero mean and a standard deviation of one. According to Ref. D-2 (p. 208) the moments of the standard normal distribution are given by

$$\mu_o = 1,$$

$$\mu_r = \begin{cases} 0, \text{ for } r \text{ odd} \\ 1 \cdot 3 \ldots (r-1) \text{ for } r \text{ even.} \end{cases}$$

A comparison of the first integral with (D-10) shows, after a little manipulation, that for N even $P_R(\underset{\sim}{X})$ is given by (D-11) plus a remainder term $E(\underset{\sim}{X})$ given by

$$E(\underset{\sim}{X}) = \frac{2C}{(2\pi A)^{N/2}\sqrt{|\underset{\approx}{M}|}} e^{\frac{B^2-AC}{2A}} \int_{\frac{B}{\sqrt{A}}}^{\infty} e^{-\frac{1}{2}W^2} \left(W - \frac{B}{\sqrt{A}}\right)^{N-1} dW \quad . \qquad \text{(D-12)}$$

By applying the binomial expansion to the integrand of (D-12) it is possible to express $E(\underset{\sim}{X})$ as a finite linear combination of incomplete gamma functions.

The simplest examples of even and odd N (except for the trivial case of N=1) are N-2 and N-3. For N=3, which was considered in Ref. D-1, (D-11) provides the earlier result

$$P_R(\underset{\sim}{X}) = \frac{e^{\frac{B^2-AC}{2A}}}{2\pi\sqrt{|\underset{\approx}{M}|}} \left(1+\frac{B^2}{A}\right) A^{-\frac{3}{2}} \quad . \qquad \text{(D-13)}$$

For N=2, (D-11) and (D-12) provide the result

$$P_R(\underset{\sim}{X}) = \frac{e^{\frac{B^2-AC}{2A}}}{\sqrt{2\pi|M|}} BA^{-\frac{3}{2}} + E(\underset{\sim}{X}) \quad , \qquad \text{(D-14)}$$

where

$$E(\underset{\sim}{X}) = \frac{e^{\frac{B^2-AC}{2A}}}{\pi A \sqrt{|\underset{\sim}{M}|}} \left(e^{-\frac{B^2}{2A}} - \frac{B}{\sqrt{A}} \int_{\frac{B}{\sqrt{A}}}^{\infty} e^{-\frac{W^2}{2}} dW \right) .$$

For cases of practical interest $C >> 1$ because the means $\bar{J}_i$ will be many standard deviations away from zero. This can be seen, for example, in the data of Ref. D-3, for which mean equivalent temperatures in the thermal bands are all of the order of 300 deg K while the standard deviations are at most 2 or 3 deg K. A similar observation can be made for the solar bands, although the means at those wavelengths (1 μ - 3 μ) differ from zero by amounts of the order of 10 standard deviations rather than 100.

If C is, in fact, large and the quantity $\frac{B}{\sqrt{A}}$ is not, the exponential factor in (D-11) and (D-12) will guarantee that $P_R(\underset{\sim}{X})$ will be negligible in general. On the other hand, when $\frac{B}{\sqrt{A}}$ is comparable to C in magnitude, i.e., when $\frac{B}{\sqrt{A}} >> 1$, it is evident from (D-12) that E(X) will be negligible. Then (D-11), which is exact when N is odd, will also provide a good approximation to $P_R(\underset{\sim}{X})$ when N is even.

2. Calculation of False Alarm Probabilities for Two-Color Systems

As observed in Chapter IV, for a two-color system the mathematical model proposed in this paper implies that in data space the decision regions determined by an optimum two-dimensional or ratio discrimination rule will always be bounded by straight lines. In fact, for a two-dimensional rule the regions

will be half planes, whereas for a ratio rule they will consist of one or more triangles or angular sectors.*

According to (3) in Chapter II, the probability of false alarm is given by

$$PFA = \frac{1}{2\pi \sqrt{|\underset{\sim}{M}_C|}} \int_R \int \exp\left[-\frac{1}{2}(\underset{\sim}{J}-\bar{\underset{\sim}{J}}_C)^t \underset{\sim}{M}_C^{-1} (\underset{\sim}{J}-\bar{\underset{\sim}{J}}_C)\right] dxdy, \quad \text{(D-15)}$$

where R is the region in which a point corresponds to a target detection as defined by the discrimination rule. To evaluate the integral in (D-15) it is convenient, first, to translate the coordinate system so that the clutter mean $\bar{\underset{\sim}{J}}_C$ is at the origin of the new system. This is done by setting

$$\underset{\sim}{r} = \underset{\sim}{J}-\underset{\sim}{J}_C \quad , \qquad \text{(D-16)}$$

whereupon (D-15) takes the form

$$PFA = \frac{1}{2\pi \sqrt{|\underset{\sim}{M}_C|}} \int_{R'} \int \exp\left[-\frac{1}{2} Q(\underset{\sim}{r})\right] dxdy \quad , \qquad \text{(D-17)}$$

where

$$Q(\underset{\sim}{r}) = \underset{\sim}{r}^t \underset{\sim}{M}_C^{-1} \underset{\sim}{r} \quad ,$$

*An angular sector may be regarded as a triangle with one side at infinity. For numerical purposes that side may have any convenient orientation, and its intersections with the other two sides of the triangle can be specified arbitrarily as long as the cartesian coordinates of the intersections and coordinate differences are large, e.g., of the order of 1000 σ.

and R' has been written in place of R as a reminder that the analytic description of the region R will be different in the new coordinate system. The next step is to change to polar coordinates; then (D-17) becomes

$$PFA = \frac{1}{2\pi \sqrt{|\underset{\approx}{M}_C|}} \int_{R'} \int \exp\left[-\frac{r^2}{2} Q(\theta)\right] r dr d\theta \ , \qquad \text{(D-18)}$$

where

$$Q(\theta) = A_{11} \cos^2\theta + 2 A_{12} \sin\theta \cos\theta + A_{22} \sin^2\theta \ . \qquad \text{(D-19)}$$

In (D-19) the coefficients A_{ij} are elements of the inverse covariance matrix $\underset{\approx}{M}_C^{-1}$ given, in terms of the standard deviations σ_1, σ_2 and the correlation coefficient ρ for clutter statistics, by

$$A_{11} = \frac{1}{(1-\rho^2)\ \sigma_1^2}, \ A_{22} = \frac{1}{(1-\rho^2)\ \sigma_2^2}, \ A_{12} = -\frac{\rho}{(1-\rho^2)\ \sigma_1 \sigma_2} . \qquad \text{(D-20)}$$

For a triangular region R' (D-18) is a sum of three terms, one for each side, of the form

$$P_i = \frac{1}{2\pi \sqrt{|\underset{\approx}{M}_C|}} \int_{\theta_i}^{\theta_{i+1}} \int_{o}^{r_i(\theta)} \exp\left[-\frac{1}{2} r^2 Q(\theta)\right] r dr d\theta \ , \qquad \text{(D-21)}$$

where θ_i and θ_{i+1} are the angular coordinates of the end-points of the side i and the equation of the line of which the side is a segment is given in polar coordinates by

$$r = r_i(\theta) = \frac{b_i}{\sin\theta - m_i \cos\theta} \ . \qquad \text{(D-22)}$$

In (D-22) b_i is the y- intercept and m_i is the slope of the line. It does not matter whether the origin of the coordinate system is inside or outside of the triangle as long as the integration over the intervals from θ_i to θ_{i+1} proceeds around the triangle in a counter-clockwise direction.

In (D-21) the integral over r can be evaluated explicitly. The result is a single integral; in fact,

$$P_i = \frac{1}{2\pi\sqrt{|\underset{\sim}{M}_C|}} \int_{\theta_i}^{\theta_{i+1}} \frac{1-\exp\left[-\frac{1}{2} r_i^2(\theta)\, Q(\theta)\right]}{Q(\theta)}\, d\theta \ , \quad \text{(D-23)}$$

where $r_i(\theta)$ is given by (D-22) and $Q(\theta)$ by (D-19).

For the 2D rule the regions R and R' are half-planes. An analysis similar to that used in deriving (D-23) leads, in this case, to the result

$$PFA = \frac{1}{2\pi\sqrt{|\underset{\sim}{M}_C|}} \int_{\tan^{-1}m}^{\tan^{-1}m+\pi} \frac{\exp\left[-\frac{1}{2} r^2(\theta)\, Q(\theta)\right]}{Q(\theta)}\, d\theta \ , \quad \text{(D-24)}$$

where

$$r(\theta) = \frac{b}{\sin\theta - m\cos\theta} \ . \quad \text{(D-25)}$$

In (D-25) b is the y- intercept and m is the slope of the line that separates the target from the background data points according to the 2D rule. The intercept b in (D-25) is defined in terms of the coordinate system centered at the mean $\underset{\sim}{\bar{J}}_C$.

The formula (D-24), for the case of a 2D rule, is in terms of a single integral that can be evaluated numerically without difficulty. For the case of a ratio rule the false alarm probability is given by

$$PFA = P_1 + P_2 + P_3 \quad , \tag{D-26}$$

wherein each term P_1 is given by a formula of the type depicted in (D-23). A straightforward numerical integration will also lead to the value of each term in (D-26).

APPENDIX D

REFERENCES

D-1. M. Burns and I. Kay, "Aspects of Signal Processing In Airborne Infrared Search and Tracking Systems," IDA Paper P-1679, September 1982 (SECRET).

D-2. H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, Princeton, New Jersey, 1946.

D-3. A.J. LaRocca and D.J. Witte, "Handbook of the Statistics of Various Terrain and Water (Ice) Backgrounds from Selected U.S. Locations," Environmental Research Institute of Michigan, Ann Arbor, Michigan, Report No. 139900-1-X, January 1980.

Note: In the preparation of this paper, no classified material from the classified references in above listing has been used.